Testing for a change of the innovation distribution in nonparametric autoregression – the sequential empirical process approach

Leonie Selk* and Natalie Neumeyer

University of Hamburg, Department of Mathematics

Bundesstrasse 55, 20146 Hamburg, Germany

January 24, 2012

Abstract

We consider a nonparametric autoregression model under conditional heteroscedasticity with the aim to test whether the innovation distribution changes in time. To this end we develop an asymptotic expansion for the sequential empirical process of nonparametrically estimated innovations (residuals). We suggest a Kolmogorov-Smirnov statistic based on the difference of the estimated innovation distributions built from the first $\lfloor ns \rfloor$ and the last $n - \lfloor ns \rfloor$ residuals, respectively ($0 \leq s \leq 1$). Weak convergence of the underlying stochastic process to a Gaussian process is proved under the null hypothesis of no change point. The result implies that the test is asymptotically distribution-free. Consistency against fixed alternatives is shown. The small sample performances of the proposed test is investigated in a simulation study and the test is applied to data examples.

1   Introduction

Assume we have observed a time series that can be modelled via an autoregression model, possibly with conditional heteroscedasticity. We aim at testing for a change point in the

---

*corresponding author; e-mail leonie.selk@math.uni-hamburg.de

innovation distribution. Tests for change points in the distribution of time series data have received a lot of attention in mathematical statistics; see Picard (1985), Giraitis, Leipus & Surgailis (1996), Horváth, Kokoszka & Teyssière (2001), Inoue (2001), Boldin (2002), Lee & Na (2004), Hušková, Prášková & Steinebach (2007), Hušková, Kirch, Prášková & Steinebach (2008), among others. Recently, an online-monitoring procedure to detect changes in the innovation distribution of linear autoregressive models was developed by Hlávka, Hušková, Kirch & Meintanis (2012). Those tests have applications in different areas, e. g. finance, climate science and medicine. For instance financial time series are tested for changes in the volatility or return (see e. g. Andreou & Ghysels (2009)) or, for climate control reasons, the annual water flow of rivers are tested for changes (see Hušková & Antoch (2003)).

Classical tests for change points in the distribution of independent data are often based on the difference of empirical distributions of the first $\lfloor ns \rfloor$ and the last $n - \lfloor ns \rfloor$ observations, respectively ($0 \leq s \leq 1$). To derive asymptotic properties of the test sequential empirical processes are considered; see Shorack & Wellner (1986, p. 131) and also Csörgö, Horváth & Szyszkowicz (1997). Those methods for independent data have been transferred to test for change points in the innovation distribution of parametric time series models. Sequential empirical processes based on estimated residuals and corresponding change point tests were suggested by Bai (1994) for ARMA-models, by Koul (1996) in the context of nonlinear time series and by Ling (1998) for nonstationary autoregressive models. Those articles are the ones most similar in spirit to the paper at hand. However, we do not assume any parametric model for either the autoregression function, nor for the conditonal variance function, but use nonparametric kernel estimation methods. The (non-sequential) empirical process of residuals in a nonparametric homoscedastic autoregressive time series model was considered by Müller, Schick & Wefelmeyer (2009) who prove an asymptotic expansion. Moreover, residual empirical processes play an important role in the test for multiplicative structure in a nonparametric heteroscedastic time series regression model by Dette, Pardo-Fernández & Van Keilegom (2009). On the other hand our approach is similar in spirit to Neumeyer & Van Keilegom (2009) who consider change point tests for the error distribution in nonparametric regression models with independent observations. However, in comparison to the latter three articles the methods of proof in the paper at hand require considerably more technical effort because both the time series structure of the data and the additional index $s \in [0, 1]$ in the stochastic process have to be taken into account.

We prove an asymptotic expansion for the sequential empirical process of residuals and prove weak convergence of the scaled and centered process to a Gaussian process. It can be seen from those results that the nonparametric estimation of the autoregression and variance function decisively changes the asymptotic behaviour in comparison to the case where innovations would be known. The asymptotic expansion of the sequential process is then used to show that nevertheless the Kolmogorov-Smirnov test for a change point as described above is asymptotically distribution-free. As a by-product of our proofs we obtain

results on uniform rates of convergence of kernel estimators (see Lemma B.1 in the appendix). Those are similar in spirit to results derived by Hansen (2008), but in contrast we avoid the stationarity assumption. Only some stabilization of the mean of innovation densities is needed (see assumption (F')), which allows us to apply the results to prove consistency of the test under the existence of a change point. We moreover present a simulation study which shows good approximations of the asymptotic level as well as good power properties of the test under the example models considered. As data applications we consider two financial time series, namely the quarterly GNP of the USA and the S&P 500 index.

The paper is organized as follows. In section 2 we present the model, the nonparametric curve estimators and the stochastic process used for the change point test. In section 3 we list technical assumptions and present the asymptotic results for the sequential empirical process as well as for the process used for the change point test under the null hypothesis of no change point. Asymptotic results under fixed alternatives are presented in section 4. Section 5 is concerned with a homoscedastic modification of the model. In section 6 we present simulation results and consider the data examples. Section 7 concludes the paper, whereas all proofs are given in the appendix.

2 Model, hypotheses and test statistic

Let $(X_j)_{j \in \mathbb{Z}}$ be a real valued stochastic process following the heteroscedastic autoregressive model of order one,

(AR) $\qquad X_j = m(X_{j-1}) + \sigma(X_{j-1})\varepsilon_j,$
where the innovations ε_j, $j \in \mathbb{Z}$, are independent with $E[\varepsilon_j] = 0$ and $E[\varepsilon_j^2] = 1 \; \forall j$ and ε_j is independent of the past X_k, $k \leq j-1$, $\forall j$.

Assume we have observed $X_0, \ldots, X_n$ and our aim is to test for a change point in the innovation distribution. Thus we formulate the null hypothesis as

$$H_0 : \qquad \varepsilon_1, \ldots, \varepsilon_n \sim F$$

(with F unknown) while the fixed alternative has the form

$$H_1 : \qquad \exists \, \theta_0 \in (0,1) : \quad \varepsilon_1, \ldots, \varepsilon_{\lfloor n\theta_0 \rfloor} \sim F, \; \varepsilon_{\lfloor n\theta_0 \rfloor + 1}, \ldots, \varepsilon_n \sim \tilde{F}, \quad F \neq \tilde{F}$$

($F, \tilde{F}$ unknown). Let $\hat{\varepsilon}_j$ denote an estimator for the innovation ε_j, $j \in \mathbb{Z}$, to be defined below. We consider a Kolmogorov-Smirnov type test statistic based on the stochastic process

$$\hat{T}_n(s,t) = \sqrt{n} \frac{\sum_{k=1}^{\lfloor ns \rfloor} w_{nk}}{n} \frac{\sum_{\ell=\lfloor ns \rfloor+1}^{n} w_{n\ell}}{n} \left(\hat{F}_{\lfloor ns \rfloor}(t) - \hat{F}^*_{n-\lfloor ns \rfloor}(t) \right), \quad s \in [0,1], t \in \mathbb{R}, \quad (2.1)$$

where the sequential empirical processes are defined as

$$\hat{F}_{\lfloor ns \rfloor}(t) = \sum_{j=1}^{\lfloor ns \rfloor} \frac{w_{nj}}{\sum_{k=1}^{\lfloor ns \rfloor} w_{nk}} I\{\hat{\varepsilon}_j \leq t\}$$

$$\hat{F}^*_{n-\lfloor ns \rfloor}(t) = \sum_{j=\lfloor ns \rfloor+1}^{n} \frac{w_{nj}}{\sum_{k=\lfloor ns \rfloor+1}^{n} w_{nk}} I\{\hat{\varepsilon}_j \leq t\}.$$

The weights are chosen as $w_{nj} = w_n(X_{j-1})$ with continuous weight function $w_n : \mathbb{R} \to [0,1]$ such that for some sequences $a_n \to -\infty$, $b_n \to \infty$,

$$w_n(x) = \begin{cases} 1, & x \in [a_n + \kappa, b_n - \kappa] \\ 0, & x \notin [a_n, b_n] \end{cases} \tag{2.2}$$

for some fixed $\kappa > 0$ independent of n. The weights are included in the definition of the sequential empirical processes to avoid problems of kernel estimation in areas where only few data are available, compare to Müller, Schick & Wefelmeyer (2009) and Dette, Pardo-Fernández & Van Keilegom (2009). Further let the residuals be defined as

$$\hat{\varepsilon}_j = \frac{X_j - \hat{m}(X_{j-1})}{\hat{\sigma}(X_{j-1})},$$

for kernel regression and variance estimators

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_{i-1}}{c_n}\right) X_i}{\sum_{i=1}^{n} K\left(\frac{x-X_{i-1}}{c_n}\right)} \tag{2.3}$$

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_{i-1}}{c_n}\right) (X_i - \hat{m}(x))^2}{\sum_{i=1}^{n} K\left(\frac{x-X_{i-1}}{c_n}\right)} \tag{2.4}$$

and $\hat{\sigma}(x) = (\hat{\sigma}^2(x))^{1/2}$. Here K denotes a kernel function and c_n a positive sequence of bandwidths. For the ease of representation we use the same bandwidth c_n to estimate m and σ, though in practice it may be advisable to choose different bandwidths. The asymptotic results presented in the paper remain valid when two different bandwidths according to the assumptions (C) and (C'), respectively, in the next sections are chosen.

We list model assumptions as well as assumptions on the estimators in the next two sections.

3   Asymptotic results under the null hypothesis

Throughout this section we make use of the following assumptions.

(K) The kernel K is a three times differentiable density with compact support $[-C, C]$ and $\sup_{u\in[-C,C]} |K^{(\mu)}(u)| \leq \bar{K} < \infty$, $\mu = 0, 1, 2, 3$. Moreover $K(C) = K(-C) = K'(C) = K'(-C) = 0$ and $\int K(u)u du = 0$.

(C) The sequence of bandwidths c_n fulfills

$$nc_n^4(\log n)^\eta \to 0, \quad \frac{(\log n)^\eta}{nc_n^{2+\sqrt{3}}} \to 0 \quad \text{for all } \eta > 0.$$

Remark: As can be seen from the proof the first bandwidth condition can be replaced by $nc_n^4(q_n q_n^\sigma)^8 (q_n^f)^2 = O(nc_n^4(\log n)^{8r_q + 8r_s + 2r_f}) = o(1)$, where $q_n, q_n^\sigma, q_n^f, r_q, r_s, r_f$ are defined in assumptions (X) and (M) below. The second bandwidth condition is equivalent to the existence of some $\delta > 0$ such that

$$\frac{(\log n)^\eta}{nc_n^{3+2\delta}} \to 0, \quad \frac{(\log n)^\eta}{nc_n^{1+\frac{1}{\delta}}} \to 0 \tag{3.1}$$

for all $\eta > 0$. The first condition in (3.1) is typical in the context of empirical processes of nonparametrically estimated residuals, compare Dette, Pardo-Fernández & Van Keilegom (2009) or Neumeyer & Van Keilegom (2009), while the log-factor stems from the boundary truncation via the weight function. The second condition in (3.1) arises at the very end of the proof of Lemma B.4 in appendix B due to a δ-dependent covering number. The constant δ is also used in Lemma B.2 in appendix B.

(I) For the interval $I_n = [a_n, b_n]$ there exists some $r_I < \infty$ such that $(b_n - a_n) = O((\log n)^{r_I})$. Moreover $\left(\int_{-\infty}^{a_n+\kappa} f_{X_0}(x)dx + \int_{b_n-\kappa}^{\infty} f_{X_0}(x)dx\right) = o((\log n)^{-1})$.

(W) The weight function $w_n : \mathbb{R} \to [0, 1]$ fulfills (2.2) and is three times differentiable such that $\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{R}} |w_n^{(\mu)}(x)| < \infty$ for $\mu = 1, 2, 3$.

(F) The innovations ε_j, $j \in \mathbb{Z}$, are identically distributed with distribution function F. Their density f is continuously differentiable and $\sup_{t\in\mathbb{R}} |f(t)t| < \infty$ as well as $\sup_{t\in\mathbb{R}} |f'(t)t^2| < \infty$.

Remark: Due to the continuity of the density f and the derivative f' it follows that also $\sup_{t\in\mathbb{R}} f(t) < \infty$, $\sup_{t\in\mathbb{R}} |f'(t)| < \infty$ and $\sup_{t\in\mathbb{R}} |f'(t)t| < \infty$.

(E) There exists some $b > 1 + \sqrt{3}$ such that $E\left[|X_0|^{2b}\right] < \infty$ and $E\left[|\varepsilon_1|^{2b}\right] < \infty$.

(X) The observations X_j, $j \in \mathbb{Z}$, are identically distributed and the process $(X_j)_{j\in\mathbb{Z}}$ is α-mixing with exponentially fast decaying mixing-coefficient $\alpha(n)$.
Their density f_{X_0} is bounded and four times differentiable with bounded derivatives. The density is also bounded away from zero on compact intervals and there exists some $r_f < \infty$ such that $q_n^f = (\inf_{x\in I_n} f_{X_0}(x))^{-1} = O((\log n)^{r_f})$.

Remark: Assumptions (F) and (X) imply strong stationarity of the process $(X_j)_{j\in\mathbb{Z}}$.

(Z) It holds that
$$\sup_{x \in J_n} \left((|m(x)| + |\sigma(x)|)^{2k} f_{X_0}(x) \right) = O(1)$$
and there exists some $1 \leq j^* < \infty$ such that
$$\sup_{x, x' \in J_n} \left((|m(x)| + |\sigma(x)|)^k (|m(x')| + |\sigma(x')|)^k f_{X_0, X_{j-1}}(x, x') \right) = O(1)$$
is valid for all $j > j^*+1$, for $k = 1, 2$, $n \to \infty$ with $J_n = [a_n - (C + c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) c_n$, $b_n + (C + c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) c_n]$.

(M) The regression function m and the scale function σ are four times differentiable and there exist some $r_q, r_s < \infty$ and q_n, q_n^σ with $q_n = O((\log n)^{r_q})$, $q_n^\sigma = O((\log n)^{r_s})$, $(q_n)^{-1} = O(1)$, $(q_n^\sigma)^{-1} = O(1)$ such that $\sup_{x \in [a_n - Cc_n, b_n + Cc_n]} |m^{(\mu)}(x)| = O(q_n)$, $\sup_{x \in [a_n - Cc_n, b_n + Cc_n]} |\sigma^{(\mu)}(x)| = O(q_n)$, $\mu = 0, 1, 2, 3, 4$ and $(\inf_{x \in I_n} |\sigma(x)|)^{-1} = O(q_n^\sigma)$.

An example for which the assumptions are fulfilled is the AR(1) model $X_j = 0.5 X_{j-1} + \varepsilon_j$ with standard normally distributed innovations ε_j, $j \in \mathbb{Z}$. Then the observations X_j, $j \in \mathbb{Z}$, are identically $\mathcal{N}(0, \frac{4}{3})$ distributed and with $I_n = [-(\frac{8}{3} \log((\log n)^2))^{1/2} - \kappa, (\frac{8}{3} \log((\log n)^2))^{1/2} + \kappa]$, a weight function that fulfills (W), a kernel function that fulfills (K) and a bandwidth that fulfills (C) all assumptions are fulfilled. To this end note that exponential α-mixing holds for stationary models with $\lim_{|x| \to \infty} (|m(x)| + |\sigma(x)| E[|\varepsilon_j|^\tau]^{\frac{1}{\tau}}) / |x| < 1$ for some $\tau \geq 1$ with $E[|\varepsilon_j|^\tau] < \infty$; see Doukhan (1994).

In the first theorem we state a stochastic expansion of the residual based sequential empirical process. The proof is given in appendix A.

Theorem 3.1 *Under model (AR) with assumptions (K), (C), (I), (W), (F), (E), (X), (Z), and (M) we have that under the null hypothesis H_0 of no change point*

$$\frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} w_{nj} \left(I\{\hat{\varepsilon}_j \leq t\} - F(t) \right)$$

$$= \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} (I\{\varepsilon_j \leq t\} - F(t)) + \frac{[ns]}{n} f(t) \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j + \frac{[ns]}{n} f(t) t \frac{1}{2n} \sum_{j=1}^{n} (\varepsilon_j^2 - 1) + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $s \in [0, 1]$ and $t \in \mathbb{R}$.

Remark 3.2 The theorem complements results by Müller, Schick & Wefelmeyer (2009) and Dette, Pardo-Fernández & Van Keilegom (2009). In both articles only non-sequential processes are considered (i. e. the case $s = 1$). While Müller, Schick & Wefelmeyer (2009) consider a homoscedastic version of model (AR) ($\sigma \equiv const$, see also section 5), Dette, Pardo-Fernández & Van Keilegom (2009) consider a heteroscedastic autoregression/regression model and a result similar to Theorem 3.1 (for $s = 1$) can be derived from their proofs. The sequential process ($s \in [0, 1]$) though requires much more involved methods of proof that also result in slightly more complicated assumptions. ∎

From the stochastic expansion weak convergence of the sequential residual process can be derived. The proof of Corollary 3.3 is given in appendix A.

Corollary 3.3 *Under the assumptions of Theorem 3.1 under the null hypothesis H_0 of no change point the process*

$$\sqrt{n}\left(\frac{\sum_{k=1}^{\lfloor ns \rfloor} w_{nk}}{n}\left(\hat{F}_{\lfloor ns \rfloor}(t) - F(t)\right)\right), \quad s \in [0,1], t \in \mathbb{R},$$

converges weakly to a centered Gaussian process $(\mathbb{K}_F(s,t))_{s \in [0,1], t \in \mathbb{R}}$ with

$$
\begin{aligned}
\mathrm{Cov}(\mathbb{K}_F(s_1,t_1), \mathbb{K}_F(s_2,t_2)) = \ &s_1 \wedge s_2 \left(F(t_1 \wedge t_2) - F(t_1)F(t_2)\right) \\
&+ s_1 s_2 \Big(f(t_1)\left(E[\varepsilon_1 I\{\varepsilon_1 \le t_2\}] + t_1 E[(\varepsilon_1^2 - 1)I\{\varepsilon_1 \le t_2\}]\right) \\
&\quad + f(t_2)\left(E[\varepsilon_1 I\{\varepsilon_1 \le t_1\}] + t_2 E[(\varepsilon_1^2 - 1)I\{\varepsilon_1 \le t_1\}]\right) \\
&\quad + f(t_1)f(t_2)\left(1 + (t_1 + t_2)E[\varepsilon_1^3] + t_1 t_2 (E[\varepsilon_1^4] - 1)\right)\Big).
\end{aligned}
$$

Remark 3.4 From Theorem 3.1 and Corollary 3.3 it can be seen that the nonparametric estimation of the autoregression and conditional variance function vastly influences the asymptotic behaviour of the process. The asymptotic distribution of the partial sum processes decicively changes when based on residuals compared to the corresponding processes built from iid innovations. This is different from simpler situations in specific parametric time series models, see Bai (1994) and Kreiß (1991), among others, but corresponds to situations in parametric as well as nonparametric regression models, see e.g. Koul (2002) and Neumeyer & Van Keilegom (2009). Note however that neither the chosen kernel function nor the bandwidth have any influence on the asymptotic distribution. ∎

The stochastic expansion given in Theorem 3.1 can be used to derive the asymptotic distribution of the change point test. First we state weak convergence of the process defined in (2.1). To this end in the following let $(\mathbb{G}(s,z))_{s \in [0,1], z \in [0,1]}$ denote a completely tucked Brownian sheet, i.e. a centered Gaussian process with covariance structure

$$\mathrm{Cov}(\mathbb{G}(s_1, z_1), \mathbb{G}(s_2, z_2)) = (s_1 \wedge s_2 - s_1 s_2)(z_1 \wedge z_2 - z_1 z_2).$$

Theorem 3.5 *Under model (AR) with the assumptions (K), (C), (I), (W), (F), (E), (X), (Z), and (M) under the null hypothesis H_0 of no change point there exist Gaussian processes $(\mathbb{G}_n(s, F(t)))_{s \in [0,1], t \in \mathbb{R}}$, $n \in \mathbb{N}$, with the same distribution as $(\mathbb{G}(s, F(t)))_{s \in [0,1], t \in \mathbb{R}}$ such that*

$$\sup_{s \in [0,1], t \in \mathbb{R}} \left|\hat{T}_n(s,t) - \mathbb{G}_n(s, F(t))\right| = o_P(1).$$

The proof is again given in appendix A as well as the proof of the next corollary in which we state the asymptotic distribution of the change point test.

Corollary 3.6 *Under the assumptions of Theorem 3.5 under the null hypothesis H_0 of no change point the Kolmogorov-Smirnov type test statistic $\sup_{s\in[0,1],t\in\mathbb{R}}|\hat{T}_n(s,t)|$ converges in distribution to $\sup_{s\in[0,1],z\in[0,1]}|\mathbb{G}(s,z)|$.*

Remark 3.7 From Corollary 3.6 it follows that the test is asymptotically distribution-free although the stochastic expansion given in Theorem 3.1 still depends on the innovation distribution in a complicated way. This remarkable feature in the context of procedures based on nonparametric residual empirical processes was already observed by Neumeyer & Van Keilegom (2009) in the context of independent observations. The critical values for the test are tabled in Picard (1985). ∎

4   Asymptotic results under fixed alternatives

The assumptions (K) and (M) as well as the following assumptions are used to proof consistency of the test under fixed alternatives.

(C') The sequence of bandwidths c_n fulfills

$$nc_n^5 \to 0, \quad \frac{(\log n)^\eta}{nc_n^{2+\sqrt{3}}} \to 0 \quad \text{for all } \eta > 0.$$

(I') For the interval $I_n = [a_n, b_n]$ there exists some $r_I < \infty$ such that $(b_n - a_n) = O((\log n)^{r_I})$.

(W') The weight function $w_n : \mathbb{R} \to [0,1]$ is continuous and fulfills (2.2).

(F') Let ε_j have distribution function F_{ε_j} and density f_{ε_j}, $j \in \mathbb{Z}$. Let $\frac{1}{n}\sum_{j=1}^n \sup_{t\in\mathbb{R}} f_{\varepsilon_j}(t) = O(1)$ as well as $\frac{1}{n}\sum_{j=1}^n \sup_{t\in\mathbb{R}} |f_{\varepsilon_j}(t)t| = O(1)$ for $n \to \infty$.

   Remark: Under the alternative H_1 the assumption is fulfilled when $\sup_{t\in\mathbb{R}} f(t) < \infty$, $\sup_{t\in\mathbb{R}} \tilde{f}(t) < \infty$, $\sup_{t\in\mathbb{R}} |f(t)t| < \infty$, and $\sup_{t\in\mathbb{R}} |\tilde{f}(t)t| < \infty$, where f and $\tilde{f}$ are densities corresponding to F and $\tilde{F}$, respectively.

(E') It holds that $\frac{1}{n}\sum_{i=1}^n E\left[|X_i|^{2b}\right] = O(1)$ for some $b > 1 + \sqrt{3}$, $n \to \infty$.

(X') The observation process $(X_j)_{j\in\mathbb{Z}}$ is α-mixing with mixing-coefficient $\alpha(n) = O(n^{-\beta})$ for some

$$\beta > \max\left(2\frac{(3+\sqrt{3})b + 2 + \sqrt{3}}{(1+\sqrt{3})b - 2(2+\sqrt{3})}, 7\right).$$

   The observation densities f_{X_i} are four times differentiable and fulfill $\sup_{x\in\mathbb{R}} n^{-1}\sum_{i=1}^n |f_{X_{i-1}}^{(\mu)}(x)| = O(1)$, $\mu = 0,1,2,3,4$. Moreover there exists some $r_f < \infty$ such that $q_n^f = (\inf_{x\in I_n} \frac{1}{n}\sum_{i=1}^n f_{X_{i-1}}(x))^{-1} = O((\log n)^{r_f})$.

(Z') For all $m_n \leq n$ with $m_n^{-1} = o(1)$ it holds that

$$\sup_{x \in J_n} \left((|m(x)| + |\sigma(x)|) \, \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} f_{X_{i-1}}(x) \right) = O(1),$$

$$\sup_{x \in J_n} \left((|m(x)| + |\sigma(x)|)^{2k} \, \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} \left(1 + E[\varepsilon_i^4]\right)^{k-1} f_{X_{i-1}}(x) \right) = O(1),$$

and there exists some $1 \leq j^* < \infty$ such that

$$\sup_{x, x' \in J_n} \left((|m(x)| + |\sigma(x)|)^k (|m(x')| + |\sigma(x')|)^k \frac{1}{m_n^2} \max_{0 \leq S \leq n - m_n} \sum_{\substack{i,j=S+1 \\ |i-j|>j^*}}^{S+m_n} f_{X_{i-1}, X_{j-1}}(x, x') \right) = O(1)$$

and

$$\sup_{x \in J_n} \left((|m(x)| + |\sigma(x)|)^{2k} \max_{j^*+1 \leq i \leq n - j^*} \sum_{j=i-j^*}^{i+j^*} \left(1 + E[\varepsilon_j^4]\right)^{k-1} f_{X_{j-1}}(x) \right) = O(1)$$

for $k = 1, 2$, $n \to \infty$ with $J_n = [a_n - (C + c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) c_n , \; b_n + (C + c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) c_n]$.

Remark: It suffices when the assumption is valid for m_n defined in (B.4) in the proof. Under the alternative H_1 the arithmetic mean of all $f_{X_{i-1}}$ converges to the weighted sum of the observation density before the change point and the long range observation density after the change point with weights θ_0 and $1 - \theta_0$, so (Z') is fulfilled if (Z) and (Z) with the long range observation density instead of f_{X_0} are fulfilled and the last part of (Z') holds.

Remark 4.1 If the observations and the innovations are identically distributed it holds that the second and third part of (X') are equivalent to the second and third part of (X) and (Z') is equivalent to (Z). The other assumptions are not equivalent, even if the the innovations are identically distributed. In detail it holds that assumption (I') is weaker than (I), (F') is weaker than (F), (E') is weaker than (E), the first part of (X') is weaker than the first part of (X), as well as (C') is weaker than (C). ∎

Remark 4.2 Note that under the alternative the process $(X_j)_{j \in \mathbb{Z}}$ is not stationary. Thus to obtain consistency most auxiliary results in appendix B are proved without assuming stationarity. A stabilisation of the density averages as in assumptions (F'), (X') is sufficient for our results to hold. In particular we generalize some of Hansen's (2008) results in Lemma B.2. ∎

Theorem 4.3 *Under the assumptions (K), (C'), (I'), (W'), (F'), (E'), (X'), (Z'), and (M) under the fixed alternative H_1 with a change point in $\lfloor n\theta_0 \rfloor$, we have*

$$\sup_{t \in \mathbb{R}} \left| \frac{\sum_{k=1}^{\lfloor n\theta_0 \rfloor} w_{nk}}{n} \left(\hat{F}_{\lfloor n\theta_0 \rfloor}(t) - F(t) \right) \right| = o_P(1)$$

$$\sup_{t \in \mathbb{R}} \left| \frac{\sum_{k=\lfloor n\theta_0 \rfloor+1}^{n} w_{nk}}{n} \left(\hat{F}^*_{n-\lfloor n\theta_0 \rfloor}(t) - \tilde{F}(t) \right) \right| = o_P(1).$$

9

Corollary 4.4 *Under the assumptions of Theorem 4.3 the Kolmogorov-Smirnov type test based on the process $\hat{T}_n$ is consistent against fixed alternatives H_1.*

The proofs of Theorem 4.3 and Corollary 4.4 are given in appendix A.

5 The homoscedastic AR-model

In this section we consider a homoscedastic AR-model

(AR1) $\qquad X_j = m(X_{j-1}) + \varepsilon_j,$
  where the innovations ε_j, $j \in \mathbb{Z}$, are independent with $E[\varepsilon_j] = 0$ and $E[\varepsilon_j^2] < \infty \ \forall j$
  and ε_j is independent of the past X_k, $k \leq j - 1$, $\forall j$.

Our aim is to test the change point hypotheses H_0 vs. H_1 from section 2. Note that here under H_1 the change in the innovation distribution can result from a change in the variance. The residuals are now defined as $\hat{\varepsilon}_j = X_j - \hat{m}(X_{j-1})$ and the test statistic is built with these in the same way as described for the heteroscedastic case; see (2.1). Let assumptions $(\bar{Z})$, $(\bar{M})$ under the null hypothesis and assumption $(Z")$ under the alternative be formulated as (Z), (M) in section 3 and (Z') in section 4, respectively, but replacing the variance function σ by a constant. Let $(\bar{F})$ be formulated as (F), but replacing conditions $\sup_{t \in \mathbb{R}} |f(t)t| < \infty$, $\sup_{t \in \mathbb{R}} |f'(t)t^2| < \infty$ by $\sup_{t \in \mathbb{R}} f(t) < \infty$, $\sup_{t \in \mathbb{R}} |f'(t)t| < \infty$. Let $(F")$ be formulated as (F'), but deleting the last condition. Let $(\bar{E})$ and $(E")$ be formulated as (E) and (E') respectively, but replacing $2b$ by b. Then the following asymptotic results are valid.

Theorem 5.1 *Under model (AR1) with assumptions (K), (C), (I), (W), $(\bar{F})$, $(\bar{E})$, (X), $(\bar{Z})$, and $(\bar{M})$ we have that under the null hypothesis H_0 of no change point*

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\left(I\{\hat{\varepsilon}_j \leq t\} - F(t)\right) = \frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor}\left(I\{\varepsilon_j \leq t\} - F(t)\right) + \frac{[ns]}{n}f(t)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $s \in [0,1]$ and $t \in \mathbb{R}$.

Corollary 5.2 *Under the assumptions of Theorem 3.1 under the null hypothesis H_0 of no change point the process*

$$\sqrt{n}\left(\frac{\sum_{k=1}^{\lfloor ns \rfloor} w_{nk}}{n}\left(\hat{F}_{\lfloor ns \rfloor}(t) - F(t)\right)\right), \quad s \in [0,1], t \in \mathbb{R},$$

converges weakly to a centered Gaussian process $(\mathbb{K}_F(s,t))_{s \in [0,1], t \in \mathbb{R}}$ with

$$\begin{aligned}
\mathrm{Cov}(\mathbb{K}_F(s_1, t_1), \mathbb{K}_F(s_2, t_2)) &= s_1 \wedge s_2 \left(F(t_1 \wedge t_2) - F(t_1)F(t_2)\right) \\
&+ s_1 s_2 \Big(f(t_1)E[\varepsilon_1 I\{\varepsilon_1 \leq t_2\}] + f(t_2)E[\varepsilon_1 I\{\varepsilon_1 \leq t_1\}] + f(t_1)f(t_2)\mathrm{Var}(\varepsilon_1)\Big).
\end{aligned}$$

Theorem 5.3 *Under model (AR1) with assumptions (K), (C'), (I'), (W'), (F"), (E"), (X'), (Z"), and ($\bar{M}$) the Kolmogorov-Smirnov type test based on the process $\hat{T}_n$ is consistent against fixed alternatives H_1.*

The proofs are analogous to the proofs of the results in sections 2 and 4, but easier due to the simpler structure of the model and the process. They are omitted for the sake of brevity.

6   Small sample performance

6.1   Simulations

The heteroscedastic model. To examine the performance of the test on small samples we considered AR(1) models and ARCH(1) models.

For the AR(1) case we considered the models

$$X_j = 0.5 \cdot X_{j-1} + \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{N}(0,1), \ \varepsilon_{\lfloor \frac{n}{2} \rfloor + 1}, \ldots, \varepsilon_n \sim \tilde{F}_1 \text{ (respectively } \tilde{F}_2),$$

where $\tilde{F}_1$ is the distribution function of a random variable that is $\mathcal{N}(-2\zeta, 1)$ distributed with probability 0.5 and $\mathcal{N}(2\zeta, 1)$ distributed with probability 0.5 and $\tilde{F}_2$ is the distribution function of a random variable that is $\mathcal{N}(0, (1-\zeta)^2)$ distributed with probability 0.5 and $\mathcal{N}(0, 2 - (1-\zeta)^2)$ distributed with probability 0.5, for different values of ζ. Though the data are generated by a homoscedastic model we assume for the data analysis validity of the heteroscedastic model (AR).

In Table 1 the rejection probabilities for 500 repetitions, level 5% and sample sizes $n \in \{100, 200\}$ are shown. They are also shown in the left panel of Figure 1. It can be seen that the level is approximated well and the power increases for increasing parameter ζ as well as for increasing sample size n.

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
$\tilde{F}_1, n = 100$	5	6.2	7	9.2	10.2	16.4	29	41
$\tilde{F}_1, n = 200$	5.8	9.8	10.6	12.4	16.2	40.2	78.2	92.4
								$\zeta = 0.99$
$\tilde{F}_2, n = 100$	5	6.8	6.4	7	7.2	10.8	18	28.2
$\tilde{F}_2, n = 200$	5.8	8.8	9.8	9.6	11.2	18.4	41.6	63.4

Table 1: *Rejection probabilities obtained from AR(1) models*

The same kind of change points was examined for ARCH(1) models

$$X_j = \sqrt{0.75 + 0.25X_{j-1}^2} \cdot \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{N}(0,1),\ \varepsilon_{\lfloor \frac{n}{2} \rfloor + 1}, \ldots, \varepsilon_n \sim \tilde{F}_1 \ (\text{respectively } \tilde{F}_2)$$

with rejection probabilities for 500 repetitions and level 5% as displayed in Table 2 and in the right panel of Figure 1.

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
$\tilde{F}_1$, $n = 100$	4.8	6.4	6.6	7.8	8	15.6	29.6	56.8
$\tilde{F}_1$, $n = 200$	5	8.4	9.2	11.2	14.6	37.2	55.4	80
								$\zeta = 0.99$
$\tilde{F}_2$, $n = 100$	4.8	6.8	7.6	8	7.4	10.2	17.6	42.2
$\tilde{F}_2$, $n = 200$	5	8.2	10	8.6	11.4	19.4	43.2	78.8

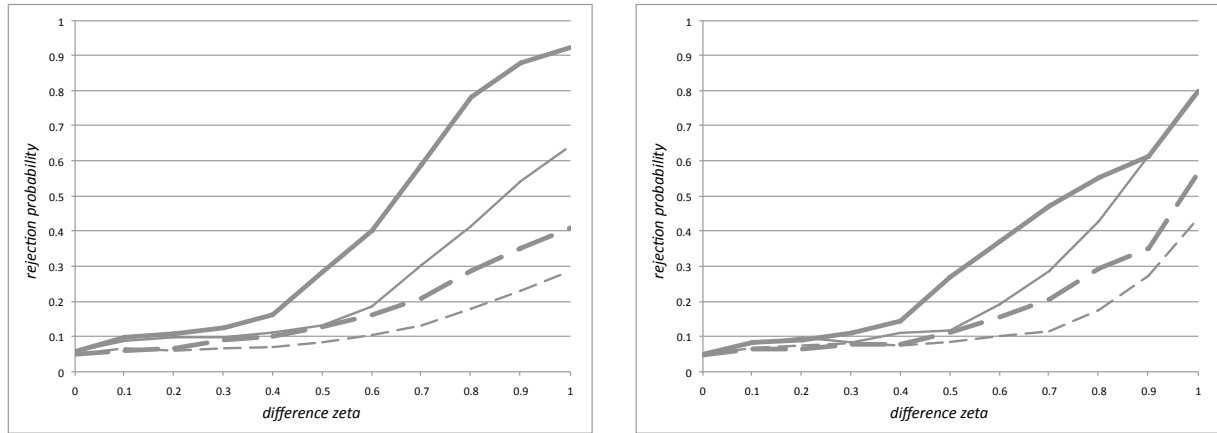Table 2: *Rejection probabilities obtained from ARCH(1) models*

Figure 1: *Rejection probabilities obtained from AR(1) (left) and ARCH(1) (right) models for $n = 100$ (dashed curves) and $n = 200$ (solid curves). The thick curves represent the results for the model with $\tilde{F}_1$, the thin curves the results for the the model with $\tilde{F}_2$.*

We also considered innovations with Student-t distribution with three degrees of freedom. The Student-t distribution has heavier tails than the normal distribution and is therefore more appropriate for modeling financial data. Due to the fact that $\text{Var}(\varepsilon_j)$ has to be one for all j, the Student-t distribution was standardized. We considered the ARCH(1) models

$$X_j = \sqrt{0.75 + 0.25X_{j-1}^2} \cdot \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{St}(3),\ \varepsilon_{\lfloor \frac{n}{2} \rfloor + 1}, \ldots, \varepsilon_n \sim \mathcal{St}(3 + 10\zeta)$$

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
$n = 100$	5.4	6.8	8.2	8.4	10.6	10	9.6	9.4
$n = 200$	6	9.6	12.8	14.6	17.2	17	18.8	19.8
$n = 500$	6.8	14.6	26.6	28.8	33	39.4	42.8	46

Table 3: *Rejection probabilities obtained from ARCH(1) models with $\mathcal{S}t(3+10\zeta)$ distributed innovations*

for different values of ζ. The rejection probabilities for 500 repetitions, level 5% and sample sizes $n \in \{100, 200, 500\}$ are shown in Table 3 and in Figure 2.

The asymptotic level is approximated reasonably well and the power increases with increasing ζ as well as with increasing n. Here the increase with ζ for small n is not as pronounced as for the models considered before, because the difference between the distribution before and after the change point is for $\zeta = 0.5$ just slightly different to that for $\zeta = 1$ because the Student-t distribution converges to the standard normal distribution.

We also examined the following ARCH(1) models:

$$X_j = \sqrt{0.75 + 0.25 X_{j-1}^2} \cdot \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{S}t(3), \ \varepsilon_{\lfloor \frac{n}{2} \rfloor + 1}, \ldots, \varepsilon_n \sim \tilde{F}_3 \ (\text{respectively } \tilde{F}_4),$$

where $\tilde{F}_3$ is the distribution function of a random variable that is $(\mathcal{S}t(3) - 2\zeta)$ distributed with probability 0.5 and $(\mathcal{S}t(3) + 2\zeta)$ distributed with probability 0.5 and $\tilde{F}_4$ is the distribution function of a random variable that is $(1 - \zeta) \cdot \mathcal{S}t(3)$ distributed with probability 0.5 and $\sqrt{2 - (1 - \zeta)^2} \cdot \mathcal{S}t(3)$ distributed with probability 0.5, for different values of ζ.

The rejection probabilities for 500 repetitions and level 5% are shown in Table 4 and Figure 3.

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
$\tilde{F}_3, n = 100$	5.4	7	7.4	12.2	20.2	32.8	34	62.4
$\tilde{F}_3, n = 200$	6	9.2	12.4	25	51.2	82	78	84.6
								$\zeta = 0.99$
$\tilde{F}_4, n = 100$	5.4	8	8.8	8.2	8.4	11.2	18.6	37.8
$\tilde{F}_4, n = 200$	6	8.6	10.2	9.2	11.6	15	39.8	77.4

Table 4: *Rejection probabilities obtained from ARCH(1) models*

The models with Student-t distributed innovations with three degrees of freedom do not fulfill the moment assumptions, because moments greater than or equal to 3 do not exist, but the simulations show that the test works on them just the same.
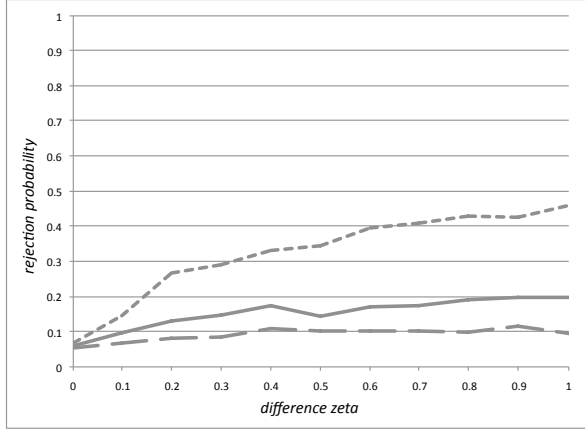
13

Figure 2: *Rejection probabilities obtained from ARCH(1) models with $\mathcal{S}t(3+10\zeta)$ distributed innovations for $n = 100$ (dashed curve), $n = 200$ (solid curve) and $n = 500$ (dotted curve).*
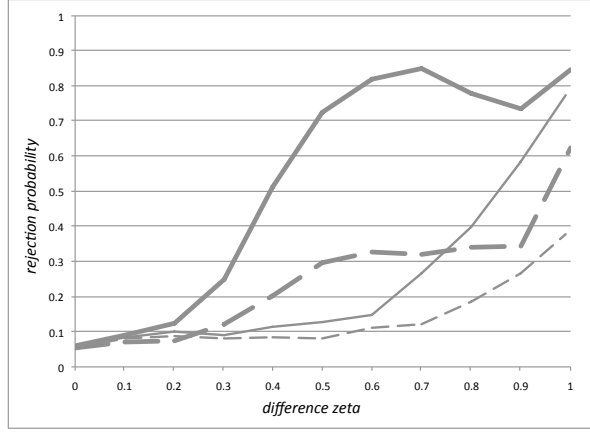
Figure 3: *Rejection probabilities obtained from ARCH(1) models for $n = 100$ (dashed curve) and $n = 200$ (solid curve). The thick curves represent the results for the model with $\tilde{F}_3$, the thin curves the results for the the model with $\tilde{F}_4$.*

Finally, we considered the skew-normal distribution as innovation distribution. Let $\tilde{F}_5$ denote the skew-normal distribution with location parameter

$$-\sqrt{\frac{2\pi\left((10\zeta)^2 + (10\zeta)^4\right)}{\pi^2 + (2\pi^2 - 2\pi)\cdot(10\zeta)^2 + (\pi^2 - 2\pi)\cdot(10\zeta)^4}},$$

scale parameter $(\pi(1 + (10\zeta)^2)^{1/2}/(\pi + (\pi - 2)(10\zeta)^2)^{1/2}$ and shape parameter 10ζ. We considered the AR(1) and ARCH(1) models

$$X_j = 0.5 \cdot X_{j-1} + \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{N}(0, 1), \ \varepsilon_{\lfloor \frac{n}{2} \rfloor+1}, \ldots, \varepsilon_n \sim \tilde{F}_5$$

and

$$X_j = \sqrt{0.75 + 0.25 X_{j-1}^2} \cdot \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{N}(0, 1), \ \varepsilon_{\lfloor \frac{n}{2} \rfloor+1}, \ldots, \varepsilon_n \sim \tilde{F}_5$$

for different values of ζ. The parameters in the skew-normal distribution were chosen like this to guarantee $E[\varepsilon_j] = 0$ and $\text{Var}(\varepsilon_j) = 1$ for all j. The rejection probabilities for 500 repetitions and level 5% are shown in Table 5 and Figure 4.

The homoscedastic model. For the homoscedastic model $X_j = m(X_{j-1}) + \varepsilon_j$ as considered in section 5 only $\text{Var}(\varepsilon_j) < \infty \ \forall j$ is assumed so that we can simulate a change in the variance. To this end we generated data from the AR(1) model

$$X_j = 0.5 \cdot X_{j-1} + \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_{\lfloor \frac{n}{2} \rfloor} \sim \mathcal{N}(0, 0.5^2), \ \varepsilon_{\lfloor \frac{n}{2} \rfloor+1}, \ldots, \varepsilon_n \sim \mathcal{N}(0, (0.5 + \zeta)^2)$$

14

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
AR(1), $n = 100$	5	7	7.2	8.6	11.4	9.8	12.2	10.8
AR(1), $n = 200$	5.8	8.4	10.2	14.8	15.2	17.2	20.8	21.6
AR(1), $n = 500$	7.8	9.2	17.2	23.4	28.8	34.8	36.8	38.8
ARCH(1), $n = 100$	4.8	7.6	8	10	11.8	11	12	12.2
ARCH(1), $n = 200$	5	7.8	11.8	12	15.2	18.8	20.6	18.6
ARCH(1), $n = 500$	7.4	9	14.8	21.8	30	35.4	36.6	38

Table 5: *Rejection probabilities obtained from AR(1) and ARCH(1) models with skew-normal distributed innovations*
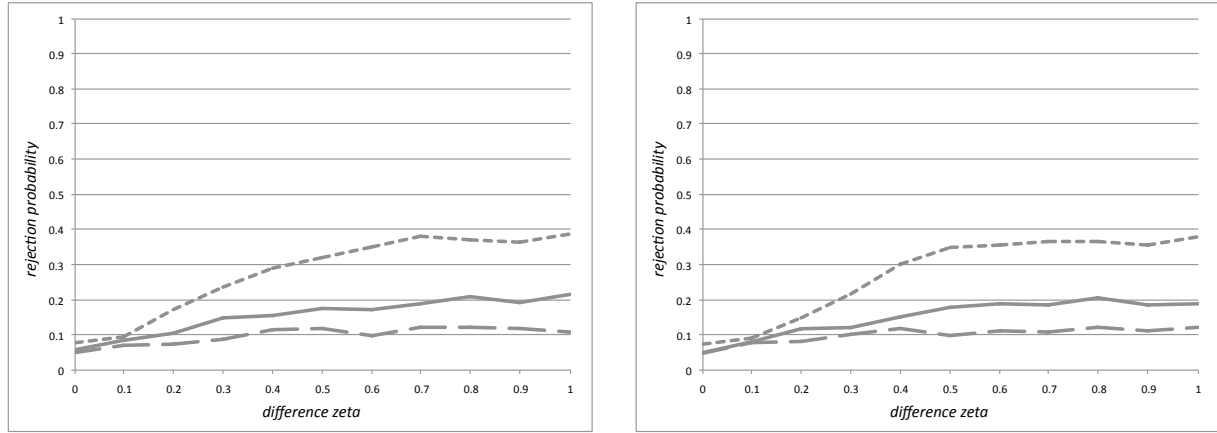
Figure 4: *Rejection probabilities obtained from AR(1) (left) and ARCH(1) (right) models with skew-normal distributed innovations for $n = 100$ (dashed curve), $n = 200$ (solid curve) and $n = 500$ (dotted curve).*

for different values of ζ.

The rejection probabilities for 500 repetitions and level 5% are shown in Table 6 and in Figure 5.

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
$n = 100$	4.6	6.8	8.6	9.8	14.4	20	29.4	39
$n = 200$	5.6	9	17.2	25.2	40.2	68.2	85.2	95.2

Table 6: *Rejection probabilities obtained from homoscedastic AR(1) models with change in variance*
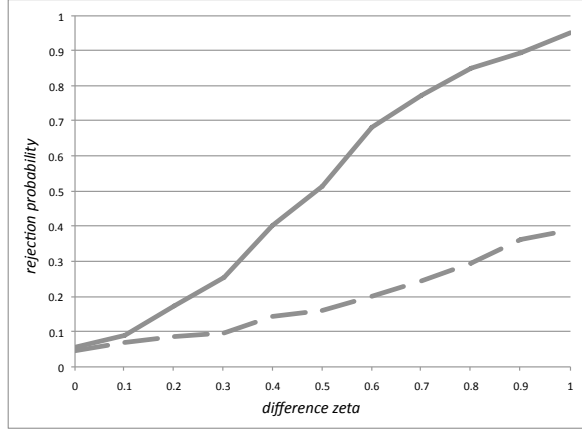
Figure 5: *Rejection probabilities obtained from homoscedastic AR(1) models with change in variance for $n = 100$ (dashed curve) and $n = 200$ (solid curve).*

It can be seen that the theoretical results are supported by the simulations.

Simulation setting. For each simulation $10 \cdot n$ observations X_j were generated, $9.5 \cdot n$ with distribution before and $0.5 \cdot n$ with distribution after the change point. For the test the last n observations were used. This was done to ensure that the process is in balance.

The empirical processes were built without the weight function w_n, which means that I_n was chosen as the real line. This is contrary to the assumptions. Nevertheless the simulations support our theoretical results very well, so it can be assumed that the weight function is necessary for the theory but the test can be used regardless.

The Nadaraya-Watson estimators $\hat{m}$ and $\hat{\sigma}$ were calculated with Gaussian kernel and bandwidth $c_n = n^{-\frac{1}{4}}$. This is also not compatible with all assumptions, e.g. the support of the kernel is not compact. However this has negligible effect on the simulations because the Gaussian kernel decreases exponentially fast at the tails. The choice of bandwidth is not compatible to the assumption as well because it does not converge faster than $n^{-\frac{1}{4}}$. A compatible choice would be $c_n = n^{-\frac{1}{4}}(\log n)^{-r}$ for some adequate $0 < r < \infty$, but for the small sample sizes that were used the logarithm would be too strong in comparison to $n^{-\frac{1}{4}}$, so we omitted it.

To study the influence of the size of bandwidth we simulated the first AR(1) model (with $\tilde{F}_1$) with $c_n = c \cdot n^{-\frac{1}{4}}$ for different values of $c \in \mathbb{R}_{>0}$. The results are shown in Figure 6. It can be seen that the rejection probability increases with c, especially for $\zeta \geq 0.5$, but also the rejection probability under the null hypothesis increases with c.
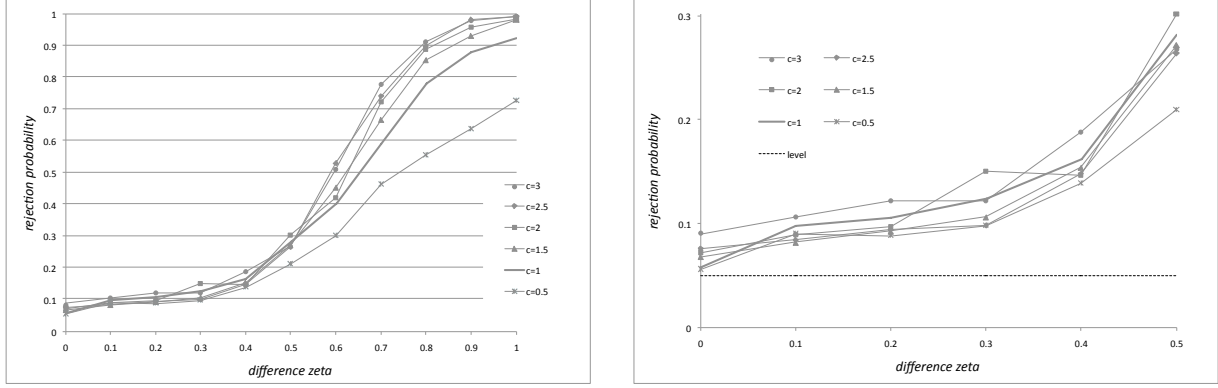
16

Figure 6: *Rejection probabilities obtained form AR(1) models with $\tilde{F}_1$ for different sizes of bandwidth and $n = 200$.*

6.2 Real data applications

We also applied our new test to real datasets. Firstly we examined the quarterly GNP (Gross National Product) of the USA in billions of dollars from 1947(1) to 2002(3). The data have been seasonally adjusted. We looked at the difference of the logarithm of the GNP, which is naturally interpreted as the growth rate of GNP. Figure 7 shows that there might be a structural break in the data and indeed our testing procedure rejects the null hypothesis of no change point with p-value smaller than 0.001. The vertical line marks the point $\lfloor ns \rfloor$ at which the test process $\hat{T}_n(s,t)$ is maximal. We used the test statistic for homoscedastic cases, next to the one for heteroscedastic cases, for these data, because the plot suggests that there might be some change in the variance. Both tests delivered the same value of the test statistic which is 1.392 (approximately).

The same data were examined in Shao & Zhang (2010) with some kind of CUSUM test that is based on an self-normalization method. They tested for a possible change in the marginal variance, 75% quantile and 25% quantile of the observations, and the test for a change in the 75% quantile rejected the null hypothesis of no change point with p-value smaller than 0.001. The tests for a change in the marginal variance and 25% quantile did not reject the null hypothesis of no change point. The p-values for these were greater than 0.1.

Shumway & Stoffer (2006) also examined these data and used stationary time series models, such as AR(1) and MA(2), to fit them. Both model fits pass their diagnostic checking tests, but our results, as well as the results of Shao & Zhang (2010), indicate that the data might not be a stable process but contain a change point.

Another dataset that we examined is the daily log-return of the S&P 500 index, a world known stock index that is quoted at the New York stock exchange, from July 1st 1998 to June 30th 2006. Figure 8 shows that there might be a structural break in these data as well, which is confirmed by our testing procedure with p-value smaller than 0.001. Again the

vertical line marks the point $\lfloor ns \rfloor$ at which the test process $\hat{T}_n(s,t)$ is maximal. Like in the GNP example we used the test statistics for both cases (hetero- and homoscedastic) and both delivered nearly the same value of the test statistic which is 1.578 for the heteroscedastic and 1.575 for the homoscedastic case (approximately).

We examined these data although it is known that for financial data often higher moments do not exist, because our simulation study with Student-t distributed innovations shows that the testing procedure works even if the moment assumption is not fulfilled.

The S&P 500 data were also examined by Kirch & Tadjuidje Kamgaing (2012). They used a testing procedure based on cumulative sums of parametrically estimated residuals for nonlinear autoregressive models. Instead of log-returns they used transformed squared log-returns and they also reject the null hypothesis of no change point. They do not give a p-value but reject clearly at level 5%.

7 Concluding remarks and outlook

In this paper we transferred classical ideas of testing for change points in samples of independent observations to testing for change points in the innovation distribution in nonparametric autoregressive models with conditional heteroscedasticity. To this end we considered the sequential empirical process of estimated innovations, proved an asymptotic expansion and weak convergence. We showed that the classical Kolmogorov-Smirnov test for a change point is asymptotically distribution-free in the new context and is not influenced asymptotically by the nonparametric estimation of the innovations. We proved consistency of the test under fixed alternatives and demonstrated the good performance in a simulation study. The proofs are based on empirical process theory for time series data and require the development of several technical auxiliary results. In particular we prove uniform rates for kernel estimators and their derivatives under nonstationarity assumptions.

It is the topic of a future project to apply the theory developed here to test for serial independence of innovations or independence of the current innovation and past observations resp. covariates in nonparametric time series regression models. Moreover our aim is to model k-dimensional joint innovation distributions in multivariate time series models.

A Proofs: main results

In this section we give the proofs for Theorems 3.1, 3.5, 4.3 and Corollaries 3.3, 3.6, 4.4, whereas some auxiliary results (Lemmata B.1–B.6) are stated and proved in section B. In some proofs standard arguments are given in condensed form for the sake of brevity. All details can be found in Selk (2011).
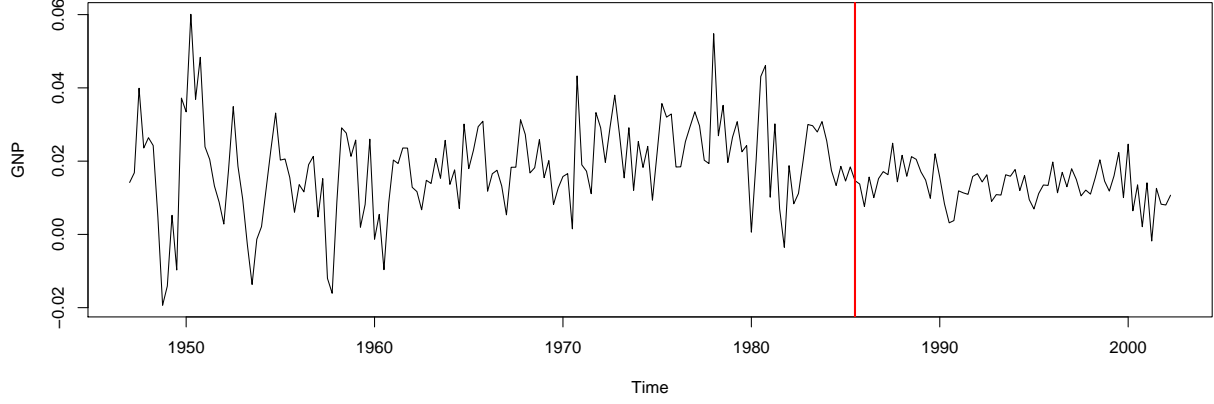
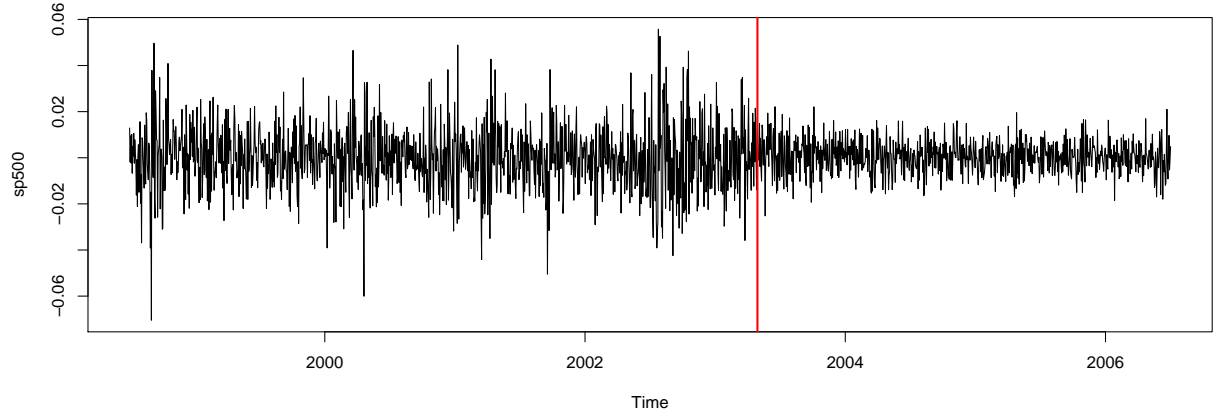Figure 7: *U.S. GNP quarterly growth rate*

Figure 8: *S&P 500 daily log-return*

A.1   Proofs for results under the null hypothesis

Proof of Theorem 3.1. From Lemma B.4 and Lemma B.5 it follows that under H_0 uniformly with respect to $s \in [0, 1]$ and $t \in \mathbb{R}$

$$\frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} w_{nj} \left(I\{\hat{\varepsilon}_j \leq t\} - F(t) \right) = \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} w_{nj} \left(I\{\varepsilon_j \leq t\} - F(t) \right) + R_n(s, t) + o_P(\frac{1}{\sqrt{n}}),$$

$$= \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} \left(I\{\varepsilon_j \leq t\} - F(t) \right) + R_n(s, t) + o_P(\frac{1}{\sqrt{n}}), \quad \text{(A.1)}$$

19

where for

$$R_n(s,t) = \frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\left(F\left(\frac{\hat{m}-m}{\sigma}(X_{j-1}) + t\frac{\hat{\sigma}}{\sigma}(X_{j-1})\right) - F(t)\right) \qquad (A.2)$$

it is straightforward to show by a first order Taylor expansion applying assumption (F) as well as Lemma B.2 that

$$R_n(s,t) = f(t)\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\left(\frac{\hat{m}-m}{\sigma}(X_{j-1}) + t\frac{\hat{\sigma}}{\sigma}(X_{j-1}) - t\right) + o_P(\frac{1}{\sqrt{n}}) \qquad (A.3)$$

uniformly with respect to s and t. Now inserting the definition of $\hat{m}$ from (2.3) we have

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\frac{\hat{m}-m}{\sigma}(X_{j-1}) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\sigma(X_{i-1})\frac{\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}K\left(\frac{X_{j-1}-X_{i-1}}{c_n}\right)\frac{1}{\sigma(X_{j-1})}}{\sum_{k=1}^{n}K\left(\frac{X_{j-1}-X_{k-1}}{c_n}\right)}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}K\left(\frac{X_{j-1}-X_{i-1}}{c_n}\right)\frac{(m(X_{i-1})-m(X_{j-1}))}{\sigma(X_{j-1})}}{\sum_{k=1}^{n}K\left(\frac{X_{j-1}-X_{k-1}}{c_n}\right)}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj} + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to s and t, which follows from Lemma B.3 (i)–(iii). Now

$$\sup_{s\in[0,1]}\left|\frac{\sum_{k=1}^{\lfloor ns \rfloor} w_{nk}}{n} - \frac{\lfloor ns \rfloor}{n}\right| = \frac{1}{n}\sum_{j=1}^{n}(1 - w_n(X_{j-1})) = o_P(1) \qquad (A.4)$$

can be shown by Chebyshev's inequality and we obtain

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\frac{\hat{m}-m}{\sigma}(X_{j-1}) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{\lfloor ns \rfloor}{n} + o_P(\frac{1}{\sqrt{n}}). \qquad (A.5)$$

By an application of

$$\frac{\hat{\sigma}}{\sigma} - 1 = \frac{\hat{\sigma}^2 - \sigma^2}{2\sigma^2} - \frac{(\hat{\sigma}-\sigma)^2}{2\sigma^2}$$

and Lemma B.2, for the second term arising from (A.3) we have

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\left(\frac{\hat{\sigma}}{\sigma}(X_{j-1}) - 1\right) = \frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\frac{\hat{\sigma}^2 - \sigma^2}{2\sigma^2}(X_{j-1}) + o_P(\frac{1}{\sqrt{n}}).$$

Now noting that

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n}K(\frac{x-X_{i-1}}{c_n})X_i^2}{\sum_{l=1}^{n}K(\frac{x-X_{l-1}}{c_n})} - (\hat{m}(x))^2 \qquad (A.6)$$

20

similarly to the derivation of (A.5) with results similar to those in Lemma B.3 one obtains

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} w_{nj}\left(\frac{\hat{\sigma}}{\sigma}(X_{j-1})-1\right) = \frac{1}{2n}\sum_{j=1}^{n}(\varepsilon_j^2-1)\frac{[ns]}{n}+o_P(\frac{1}{\sqrt{n}}). \tag{A.7}$$

Finally from (A.3), (A.5) and (A.7) one has

$$R_n(s,t) = \frac{[ns]}{n}f(t)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j + \frac{[ns]}{n}f(t)t\frac{1}{2n}\sum_{j=1}^{n}(\varepsilon_j^2-1)+o_P(\frac{1}{\sqrt{n}}).$$

and the assertion follows from this equality and (A.1). □

Proof of Corollary 3.3. By an application of Theorem 3.1 it remains to show that

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor}(I\{\varepsilon_j \le t\}-F(t)) + \frac{\lfloor ns \rfloor}{n}f(t)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j + \frac{\lfloor ns \rfloor}{n}f(t)t\frac{1}{2n}\sum_{j=1}^{n}(\varepsilon_j^2-1)\right)$$

$$= \sum_{j=1}^{n}\big(Z_{nj}(s,t,f(t),f(t)t)-E\left[Z_{nj}(s,t,f(t),f(t)t)\right]\big) \tag{A.8}$$

converges weakly to the process $(\mathbb{K}_F(s,t))_{s\in[0,1],t\in\mathbb{R}}$. Here we use the notations

$$Z_{nj}(s,t,u,v) = n^{-\frac{1}{2}}\left(I\{\varepsilon_j \le t\}I\left\{\frac{j}{n}\le s\right\} + \frac{\lfloor ns \rfloor}{n}u\varepsilon_j + \frac{\lfloor ns \rfloor}{n}\frac{1}{2}v(\varepsilon_j^2-1)\right), \quad (s,t,u,v)\in\mathcal{F},$$

where

$$\mathcal{F}=\left\{(s,t,u,v):\ s\in[0,1],t\in\mathbb{R},u\in\left[0,\sup_{t\in\mathbb{R}}f(t)\right],v\in\left[-\sup_{t\in\mathbb{R}}|f(t)t|,\sup_{t\in\mathbb{R}}|f(t)t|\right]\right\}$$

is equipped with the semi-metric

$$\rho((s,t,u,v),(s',t',u',v')) = |s-s'|+|F(t)-F(t')|+|u-u'|+|v-v'|$$

and is totally bounded. By an application of Theorem 2.11.9 in van der Vaart & Wellner (1996) one can show weak convergence of the process

$$\left(\sum_{j=1}^{n}\big(Z_{nj}(s,t,u,v)-E\left[Z_{nj}(s,t,u,v)\right]\big)\right)_{(s,t,u,v)\in\mathcal{F}}$$

to a centered Gaussian process. Details are omitted for the sake of brevity, but the arguments are similar to (but simpler than) those in the proofs in Neumeyer & Van Keilegom (2009) (see their proof of theorem 3 and the online supporting information).

The assertion now follows by the continuous mapping theorem applied to the projections $u=f(t)$ and $v=f(t)t$ and by a straightforward calculation of the asymptotic covariance. □

Proof of Theorem 3.5. For the process $\hat{T}_n$ defined in (2.1) we have by a straightforward calculation

$$
\begin{aligned}
&\hat{T}_n(s,t) \\
&= \sqrt{n}\left(\frac{\sum_{k=\lfloor ns\rfloor+1}^{n} w_{nk}}{n}\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} w_{nj}I\{\hat{\varepsilon}_j \le t\} - \frac{\sum_{k=1}^{\lfloor ns\rfloor} w_{nk}}{n}\frac{1}{n}\sum_{j=\lfloor ns\rfloor+1}^{n} w_{nj}I\{\hat{\varepsilon}_j \le t\}\right) \\
&= \sqrt{n}\left(\frac{\sum_{k=1}^{n} w_{nk}}{n}\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} w_{nj}\left(I\{\hat{\varepsilon}_j \le t\} - F(t)\right) - \frac{\sum_{k=1}^{\lfloor ns\rfloor} w_{nk}}{n}\frac{1}{n}\sum_{j=1}^{n} w_{nj}\left(I\{\hat{\varepsilon}_j \le t\} - F(t)\right)\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} w_{nj}\left(I\{\hat{\varepsilon}_j \le t\} - F(t)\right) - \frac{\lfloor ns\rfloor}{n}\frac{1}{n}\sum_{j=1}^{n} w_{nj}\left(I\{\hat{\varepsilon}_j \le t\} - F(t)\right)\right) + o_P(1)
\end{aligned}
$$

uniformly with respect to s and t. Here the last equality follows from (A.4). Now inserting the expansion given in Theorem 3.1 we directly obtain

$$
\begin{aligned}
&\hat{T}_n(s,t) \\
&= \sqrt{n}\frac{\lfloor ns\rfloor}{n}\left(1 - \frac{\lfloor ns\rfloor}{n}\right)\left(\frac{1}{\lfloor ns\rfloor}\sum_{j=1}^{\lfloor ns\rfloor} I\{\varepsilon_j \le t\} - \frac{1}{n-\lfloor ns\rfloor}\sum_{j=\lfloor ns\rfloor+1}^{n} I\{\varepsilon_j \le t\}\right) + o_P(1)
\end{aligned}
$$

uniformly with respect to s and t. The assertion follows from Remark 2 in Neumeyer & Van Keilegom (2009), which goes back to Theorem 3.1 by Csörgő, Horváth & Szyszkowicz (1997). □

Proof of Corollary 3.6. It directly follows from Theorem 3.5 and the continuous mapping theorem that

$$
\sup_{s\in[0,1],t\in\mathbb{R}}\left|\hat{T}_n(s,t)\right| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{s\in[0,1],t\in\mathbb{R}}|\mathbb{G}(s,F(t))| = \sup_{s\in[0,1],z\in[0,1]}|\mathbb{G}(s,z)|
$$

where the last equality holds by continuity of F. □

A.2 Proofs for results under fixed alternatives

Proof of Theorem 4.3. Assume that H_1 is valid with a change point in $\lfloor n\theta_0\rfloor$. Analogously to the proof of Theorem 3.1 we have from Lemma B.4 that uniformly with respect to $t \in \mathbb{R}$

$$
\begin{aligned}
\frac{\sum_{k=1}^{\lfloor n\theta_0\rfloor} w_{nk}}{n}\left(\hat{F}_{\lfloor n\theta_0\rfloor}(t) - F(t)\right) &= \frac{1}{n}\sum_{j=1}^{\lfloor n\theta_0\rfloor} w_{nj}(I\{\hat{\varepsilon}_j \le t\} - F(t)) \\
&= \frac{1}{n}\sum_{j=1}^{\lfloor n\theta_0\rfloor} w_{nj}\left(I\{\varepsilon_j \le t\} - F(t)\right) + R_n(\theta_0,t) + o_P(1) \\
&= R_n(\theta_0,t) + o_P(1),
\end{aligned}
$$

22

where the last equality follows from Lemma B.6. Here, for R_n as defined in (A.2) one has by the mean value theorem applying assumption (F') and Lemma B.2 that

$$\sup_{t\in\mathbb{R}}|R_n(\theta_0,t)| \le \sup_{x\in I_n}\left|\frac{\hat{m}-m}{\sigma}(x)\right|\sup_{t\in\mathbb{R}}|f(t)| + \sup_{x\in I_n}\left|\frac{\hat{\sigma}-\sigma}{\sigma}(x)\right|\sup_{t\in\mathbb{R}}|f(t)t|O_P(1) = o_P(1).$$

Thus the first assertion of the Theorem follows. The second assertion is shown analogously. $\square$

Proof of Corollary 4.4. Assume that H_1 is valid with a change point in $\lfloor n\theta_0\rfloor$. For $\hat{T}_n$ defined in (2.1) we have

$$\sup_{t\in\mathbb{R}}\left|\hat{T}_n(\theta_0,t)\right| = \sqrt{n}\sup_{t\in\mathbb{R}}\left|\frac{\sum_{k=1}^{\lfloor n\theta_0\rfloor}w_{nk}}{n}\frac{\sum_{k=\lfloor n\theta_0\rfloor+1}^{n}w_{nk}}{n}\left(\hat{F}_{\lfloor n\theta_0\rfloor}(t) - \hat{F}^*_{n-\lfloor n\theta_0\rfloor}(t)\right)\right|$$

$$\ge \sqrt{n}\left(\sup_{t\in\mathbb{R}}\left|\frac{\sum_{k=1}^{\lfloor n\theta_0\rfloor}w_{nk}}{n}\frac{\sum_{k=\lfloor n\theta_0\rfloor+1}^{n}w_{nk}}{n}\left(F(t) - \tilde{F}(t)\right)\right| - o_P(1)\right)$$

from Theorem 4.3. Now under H_1 the right hand side converges to infinity in probability from which consistency follows. $\square$

B   Proofs: auxiliary results

B.1   Results

For easy overview we first state the auxiliary results and then collect the proofs in the next subsection.

Lemma B.1 *Under the assumptions of either Theorem 3.1 or Theorem 4.3 we have that*

$$\sup_{x\in I_n}\left|\frac{1}{nc_n}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{x-X_{i-1}}{c_n}\right)X_i^k - E\left[\frac{1}{nc_n}\sum_{i=1}^{n}K^{(\nu)}\left(\frac{x-X_{i-1}}{c_n}\right)X_i^k\right]\right| = O_P(\bar{\epsilon}_n)$$

for $\nu = 0,1,2$, $k = 0,1,2$, where $\bar{\epsilon}_n = c_n^{-\frac{1}{2}}n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$.

Lemma B.2 *Under the assumptions of either Theorem 3.1 or Theorem 4.3 we have that*

(i) $\sup_{x\in I_n}\left|\frac{\hat{m}(x)-m(x)}{\sigma(x)}\right| = O_P\left(\left(c_n^{-\frac{1}{2}}n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} + c_n^2\right)q_nq_n^fq_n^\sigma\right) = o_P(1),$

$\sup_{x\in I_n}\left|\frac{\hat{\sigma}(x)-\sigma(x)}{\sigma(x)}\right| = O_P\left(\left(c_n^{-\frac{1}{2}}n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} + c_n^2\right)(q_nq_n^fq_n^\sigma)^2\right) = o_P(1),$ *where q_n, q_n^f, q_n^σ are defined in (M) and (X) resp. (X').*

(ii) $\sup_{x\in I_n}\left|\frac{\partial}{\partial x}\left(\frac{\hat{m}(x)-m(x)}{\sigma(x)}\right)\right| = o_P(1), \sup_{x\in I_n}\left|\frac{\partial}{\partial x}\left(\frac{\hat{\sigma}(x)-\sigma(x)}{\sigma(x)}\right)\right| = o_P(1)$

23

(iii) $\displaystyle\sup_{x,y\in I_n, x\neq y}\frac{\left|\frac{\partial}{\partial x}\left(\frac{\hat{m}(x)-m(x)}{\sigma(x)}\right)-\frac{\partial}{\partial y}\left(\frac{\hat{m}(y)-m(y)}{\sigma(y)}\right)\right|}{|y-x|^\delta}=o_P(1),$

$\displaystyle\sup_{x,y\in I_n, x\neq y}\frac{\left|\frac{\partial}{\partial x}\left(\frac{\hat{\sigma}(x)-\sigma(x)}{\sigma(x)}\right)-\frac{\partial}{\partial y}\left(\frac{\hat{\sigma}(y)-\sigma(y)}{\sigma(y)}\right)\right|}{|y-x|^\delta}=o_P(1)\ \text{with }\delta\text{ from (3.1).}$

Lemma B.3 *Under the assumptions of Theorem 3.1 we have*

(i) $\displaystyle\sup_{s\in[0,1]}\left|\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon_i\sum_{j=1}^{\lfloor ns\rfloor}w_{nj}\left(\frac{\frac{1}{c_n}K\left(\frac{X_{j-1}-X_{i-1}}{c_n}\right)-\frac{1}{nc_n}\sum_{k=1}^{n}K\left(\frac{X_{j-1}-X_{k-1}}{c_n}\right)}{\frac{1}{nc_n}\sum_{k=1}^{n}K\left(\frac{X_{j-1}-X_{k-1}}{c_n}\right)}\right)\right|=o_P\left(\frac{1}{\sqrt{n}}\right)$

(ii) $\displaystyle\sup_{s\in[0,1]}\left|\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon_i\sum_{j=1}^{\lfloor ns\rfloor}\frac{w_{nj}}{\sigma(X_{j-1})}\left(\frac{\frac{1}{c_n}K\left(\frac{X_{j-1}-X_{i-1}}{c_n}\right)(\sigma(X_{i-1})-\sigma(X_{j-1}))}{\frac{1}{nc_n}\sum_{k=1}^{n}K\left(\frac{X_{j-1}-X_{k-1}}{c_n}\right)}\right)\right|=o_P\left(\frac{1}{\sqrt{n}}\right)$

(iii) $\displaystyle\sup_{s\in[0,1]}\left|\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor}\frac{w_{nj}}{\sigma(X_{j-1})}\left(\frac{\frac{1}{nc_n}\sum_{i=1}^{n}K\left(\frac{X_{j-1}-X_{i-1}}{c_n}\right)(m(X_{i-1})-m(X_{j-1}))}{\frac{1}{nc_n}\sum_{k=1}^{n}K\left(\frac{X_{j-1}-X_{k-1}}{c_n}\right)}\right)\right|=o_P\left(\frac{1}{\sqrt{n}}\right).$

Lemma B.4 *Under the assumptions of either Theorem 3.1 or Theorem 4.3 we have that*

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor}w_{nj}\left(I\{\hat{\varepsilon}_j\leq t\}-F_{\varepsilon_j}\left(t\frac{\hat{\sigma}}{\sigma}(X_{j-1})+\frac{\hat{m}-m}{\sigma}(X_{j-1})\right)-I\{\varepsilon_j\leq t\}+F_{\varepsilon_j}(t)\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly with respect to $s\in[0,1]$ and $t\in\mathbb{R}$, where under H_0 all F_{ε_j} are equal to F.

Lemma B.5 *Under the assumptions of Theorem 3.1 we have that*

$$\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor}w_{nj}\left(I\{\varepsilon_j\leq t\}-F(t)\right)=\frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor}\left(I\{\varepsilon_j\leq t\}-F(t)\right)+o_P\left(\frac{1}{\sqrt{n}}\right)$$

uniformly with respect to $s\in[0,1]$ and $t\in\mathbb{R}$.

Lemma B.6 *Under the assumptions of Theorem 4.3 we have that*

$$\frac{1}{n}\sum_{j=1}^{\lfloor n\theta_0\rfloor}w_{nj}\left(I\{\varepsilon_j\leq t\}-F(t)\right)=o_P(1),\quad\frac{1}{n}\sum_{j=\lfloor n\theta_0\rfloor+1}^{n}w_{nj}\left(I\{\varepsilon_j\leq t\}-\tilde{F}(t)\right)=o_P(1)$$

uniformly with respect to $t\in\mathbb{R}$.

B.2 Proofs

Proof of Lemma B.1. Let $k \in \{0, 1, 2\}$. Throughout the proof we assume that $|X_i^k| \leq n^{\frac{1}{b}} \log n$ for all i with b from assumptions (E), (E'). This is possible, because

$$P \left(\max_{1 \leq i \leq n} |X_i^k| > n^{\frac{1}{b}} \log n \right) \leq \sum_{i=1}^n P \left(|X_i^k| > n^{\frac{1}{b}} \log n \right) \leq \frac{1}{n(\log n)^b} \sum_{i=1}^n E \left[|X_i|^{kb} \right] = o(1)$$

by assumption (E) resp. (E').

Choose points x_j, $j = 1, \ldots, M_n^* \leq M_n = (b_n - a_n)/(\bar{\epsilon}_n c_n)$ (for $I_n = [a_n, b_n]$ from assumption (I) resp. (I')) such that I_n is covered by intervals $[x_j - \bar{\epsilon}_n c_n, x_j + \bar{\epsilon}_n c_n]$. Now let $\bar{K}_I = \bar{K} I_{[-C,C]}$, where $K^{(\nu+1)}$ is bounded by $\bar{K}$ on the support $[-C, C]$ of K for $\nu \in \{0, 1, 2\}$ (assumption (K)). Then by the mean value theorem we have

$$\left| \frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x - X_{i-1}}{c_n} \right) X_i^k - E \left[\frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x - X_{i-1}}{c_n} \right) X_i^k \right] \right|$$

$$\leq \left| \frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x_j - X_{i-1}}{c_n} \right) X_i^k - E \left[\frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x_j - X_{i-1}}{c_n} \right) X_i^k \right] \right|$$

$$+ \bar{\epsilon}_n \frac{1}{nc_n} \sum_{i=1}^n \bar{K}_I \left(\frac{x_j - X_{i-1}}{c_n} \right) |X_i^k| + \bar{\epsilon}_n E \left[\frac{1}{nc_n} \sum_{i=1}^n \bar{K}_I \left(\frac{x_j - X_{i-1}}{c_n} \right) |X_i^k| \right]$$

for all x such that $|x - x_j| \leq \bar{\epsilon}_n c_n$. From this we obtain

$$\sup_{x \in I_n} \left| \frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x - X_{i-1}}{c_n} \right) X_i^k - E \left[\frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x - X_{i-1}}{c_n} \right) X_i^k \right] \right|$$

$$\leq \max_{1 \leq j \leq M_n^*} \left| \frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x_j - X_{i-1}}{c_n} \right) X_i^k - E \left[\frac{1}{nc_n} \sum_{i=1}^n K^{(\nu)} \left(\frac{x_j - X_{i-1}}{c_n} \right) X_i^k \right] \right| \quad \text{(B.1)}$$

$$+ \max_{1 \leq j \leq M_n^*} \bar{\epsilon}_n \left| \frac{1}{nc_n} \sum_{i=1}^n \bar{K}_I \left(\frac{x_j - X_{i-1}}{c_n} \right) |X_i^k| - E \left[\frac{1}{nc_n} \sum_{i=1}^n \bar{K}_I \left(\frac{x_j - X_{i-1}}{c_n} \right) |X_i^k| \right] \right| \quad \text{(B.2)}$$

$$+ 2\bar{\epsilon}_n \max_{1 \leq j \leq M_n^*} E \left[\frac{1}{nc_n} \sum_{i=1}^n \bar{K}_I \left(\frac{x_j - X_{i-1}}{c_n} \right) |X_i^k| \right].$$

The last term on the right hand side can be bounded by

$$2\bar{\epsilon}_n \sup_{x \in I_n} E \left[\frac{1}{nc_n} \sum_{i=1}^n \bar{K}_I \left(\frac{x - X_{i-1}}{c_n} \right) |X_i^k| \right]$$

$$= 2\bar{\epsilon}_n \sup_{x \in I_n} \frac{1}{n} \sum_{i=1}^n \int \bar{K}_I(u) E \left[|X_i^k| \, \Big| \, X_{i-1} = x - uc_n \right] f_{X_{i-1}}(x - uc_n) \, du$$

$$\leq 2\bar{\epsilon}_n \int \bar{K}_I(u) \, du \sup_{x \in I_n} \frac{1}{n} \sum_{i=1}^n E \left[|X_i^k| \, \Big| \, X_{i-1} = x \right] f_{X_{i-1}}(x) = O(\bar{\epsilon}_n)$$

25

by a change of variable and by assumption (K) and (Z) or (Z'), respectively, and model assumption (AR). In what follows we show that term (B.1) is of order $O_P(\bar{\epsilon}_n)$; it can be shown analogously that (B.2) is of the same order. Define (for j fixed)

$$Y_i = K^{(\nu)}\left(\frac{x_j - X_{i-1}}{c_n}\right) X_i^k - E\left[K^{(\nu)}\left(\frac{x_j - X_{i-1}}{c_n}\right) X_i^k\right],$$

then the sequence $(Y_i)_i$ inherits the mixing conditions from $(X_i)_i$ due to 2.6.1 (ii) in Fan & Yao (2005). Further the variables are centered and bounded by $2\bar{K}n^{1/b}\log n$. We apply Liebscher's (1996) Theorem 2.1 to $\sum_{i=1}^n Y_i$ to obtain

$$P\left(\max_{1 \leq j \leq M_n^*}\left|\frac{1}{nc_n}\sum_{i=1}^n K^{(\nu)}\left(\frac{x_j - X_{i-1}}{c_n}\right) X_i^k - E\left[\frac{1}{nc_n}\sum_{i=1}^n K^{(\nu)}\left(\frac{x_j - X_{i-1}}{c_n}\right) X_i^k\right]\right| > \bar{M}\bar{\epsilon}_n\right)$$

$$\leq M_n\left(4\exp\left(-\frac{nc_n\bar{M}^2\bar{\epsilon}_n^2}{64\left(1 + m_n c_n\right) A(m_n) + \frac{16}{3}\bar{M}\bar{K}m_n n^{\frac{1}{b}}\bar{\epsilon}_n \log(n)}\right) + 4\frac{n}{m_n}\alpha(m_n)\right) \qquad \text{(B.3)}$$

for some $\bar{M}$ independent of j and for

$$m_n = \begin{cases} \lfloor n\bar{\epsilon}_n^2(\log M_n)^{-1}(\log n)^{-1}\rfloor & \text{if } n^{\frac{1}{b}-\frac{1}{2}}c_n^{-\frac{3}{2}}(\log n)^{\frac{5}{2}} = O(1) \\ \lfloor n^{1-\frac{1}{b}}c_n\bar{\epsilon}_n(\log M_n)^{-1}(\log n)^{-2}\rfloor & \text{otherwise.} \end{cases} \qquad \text{(B.4)}$$

Further,

$$A(m_n) = 2(C + \bar{\epsilon}_n)\bar{K}^2 \sup_{x \in J_n} \max_{j^*+1 \leq i \leq n-j^*} \sum_{j=i-j^*}^{i+j^*} \sqrt{E\left[X_j^{2k}\Big|X_{j-1} = x\right] f_{X_{j-1}}(x)}$$

$$\times \sup_{x \in J_n}\frac{1}{m_n}\max_{0 \leq S \leq n-m_n}\sum_{i=S+1}^{S+m_n}\sqrt{E\left[X_i^{2k}\Big|X_{i-1} = x\right] f_{X_{i-1}}(x)}$$

$$+ \left(2(C + \bar{\epsilon}_n)\bar{K}\right)^2\left(\left(\sup_{x \in J_n}\frac{1}{m_n}\max_{0 \leq S \leq n-m_n}\sum_{i=S+1}^{S+m_n} E\left[|X_i|^k\Big|X_{i-1} = x\right] f_{X_{i-1}}(x)\right)^2\right.$$

$$\left.+ \sup_{x,x' \in J_n}\frac{1}{m_n^2}\max_{0 \leq S \leq n-m_n}\sum_{\substack{i,j=S+1 \\ |i-j|>j^*}}^{S+m_n} E\left[|X_i|^k|X_j|^k\Big|X_{i-1} = x, X_{j-1} = x'\right] f_{X_{i-1},X_{j-1}}(x,x')\right),$$

with j^* as in assumption (Z) resp. (Z'). In order to obtain B.3 from Liebscher's Theorem one has to show that

$$\max_{0 \leq T \leq n-1} E\left[\left(\sum_{i=T+1}^{\min(T+m_n,n)} Y_i\right)^2\right] \leq \left(m_n c_n + m_n^2 c_n^2\right) A(m_n).$$

This can be done by some tedious calculations, which are omitted for the sake of brevity.

By assumption (Z) resp. (Z') we have $A(m_n) = O(1)$. To see this consider for example for $k = 1$ the term

$$\sup_{x \in J_n} \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} E\left[|X_i| \,\Big|\, X_{i-1} = x\right] f_{X_{i-1}}(x)$$

$$= \sup_{x \in J_n} \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} E\left[|m(x) + \sigma(x)\varepsilon_i|\right] f_{X_{i-1}}(x)$$

$$\leq \sup_{x \in J_n} \left((|m(x)| + 2|\sigma(x)|) \, \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} f_{X_{i-1}}(x) \right) = O(1)$$

(note that $m_n^{-1} = o(1)$), and

$$\sup_{x \in J_n} \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} \sqrt{E\left[X_i^2 \,\Big|\, X_{i-1} = x\right] f_{X_{i-1}}(x)}$$

$$\leq \sup_{x \in J_n} \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \left(m_n + \sum_{\substack{i=S+1 \\ E\left[X_i^2 | X_{i-1}=x\right] f_{X_{i-1}}(x) > 1}}^{S+m_n} E\left[X_i^2 \,\Big|\, X_{i-1} = x\right] f_{X_{i-1}}(x) \right)$$

$$\leq 1 + \sup_{x \in J_n} \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} E\left[X_i^2 \,\Big|\, X_{i-1} = x\right] f_{X_{i-1}}(x)$$

$$\leq 1 + \sup_{x \in J_n} \left((|m(x)| + |\sigma(x)|)^2 \, \frac{1}{m_n} \max_{0 \leq S \leq n - m_n} \sum_{i=S+1}^{S+m_n} f_{X_{i-1}}(x) \right) = O(1).$$

The other terms in the definition of $A(m_n)$ are treated similarly.

Inserting the definitions of M_n, $\bar{\epsilon}_n$, m_n and $A(m_n) = O(1)$ one obtains with a simply calculation that (B.3) is of order $o(1)$ by the assumptions on the bandwidth c_n and the mixing coefficient. This concludes the proof. $\qquad\square$

Proof of Lemma B.2. We only present the proofs for the assertions on $\hat{m}$, those on $\hat{\sigma}$ follow by similar arguments using (A.6).

Let $\hat{g}_k(x) = \frac{1}{nc_n} \sum_{i=1}^{n} K\left(\frac{x - X_{i-1}}{c_n}\right) X_i^k$ for $k \in \{0, 1, 2\}$. Then from Lemma B.1 it directly follows that

$$\sup_{x \in I_n} \left| \frac{\partial^\nu}{\partial x^\nu} \left(\hat{g}_k(x) - E\left[\hat{g}_k(x)\right] \right) \right| = O_P\left(\frac{\bar{\epsilon}_n}{c_n^\nu} \right) = O_P\left(c_n^{-\frac{1}{2} - \nu} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} \right). \qquad (B.5)$$

27

Let further $g_{k,i}(x) = E\left[X_i^k | X_{i-1} = x\right] f_{X_{i-1}}(x)$ for $i = 1, \ldots, n$. Then

$$\sup_{x \in I_n} \left| \frac{\partial^\nu}{\partial x^\nu} \left(E\left[\hat{g}_k(x)\right] - \frac{1}{n} \sum_{i=1}^n g_{k,i}(x) \right) \right|$$

$$= \sup_{x \in I_n} \left| \frac{1}{c_n^\nu} \frac{1}{nc_n} \sum_{i=1}^n \int K^{(\nu)} \left(\frac{x-y}{c_n} \right) E\left[X_i^k \Big| X_{i-1} = y\right] f_{X_{i-1}}(y) dy - \frac{1}{n} \sum_{i=1}^n g_{k,i}^{(\nu)}(x) \right|$$

$$= \sup_{x \in I_n} \left| \frac{1}{c_n^\nu} \frac{1}{n} \sum_{i=1}^n \int K^{(\nu)}(u) \, g_{k,i}(x - uc_n) du - \frac{1}{n} \sum_{i=1}^n g_{k,i}^{(\nu)}(x) \right|.$$

Noting that by assumption (K),

$$\int K^{(\nu)}(u) \, u^\mu du = 0 \text{ for } \mu \leq \nu + 1, \ \mu \neq \nu, \quad \int K^{(\nu)}(u) \, u^\nu du = (-1)^\nu \nu!$$

for $\nu \in \{0, 1, 2\}$, from a Taylor expansion it further follows that

$$\sup_{x \in I_n} \left| \frac{\partial^\nu}{\partial x^\nu} \left(E\left[\hat{g}_k(x)\right] - \frac{1}{n} \sum_{i=1}^n g_{k,i}(x) \right) \right|$$

$$\leq \frac{c_n^2}{(\nu+2)!} \sup_{x \in J_n} \left| \frac{1}{n} \sum_{i=1}^n g_{k,i}^{(\nu+2)}(x) \right| \left| \int K^{(\nu)}(u) \, u^{\nu+2} du \right| = O(c_n^2 q_n^k), \quad \text{(B.6)}$$

where the last equality follows from assumption (X) resp. (X') and (M) (note that $g_{0,i} = f_{X_{i-1}}$, $g_{1,i} = m f_{X_{i-1}}$, $g_{2,i} = (m^2 + \sigma^2) f_{X_{i-1}}$). Note further that

$$\inf_{x \in I_n} \left| \frac{\hat{g}_0(x)}{\frac{1}{n} \sum_{i=1}^n g_{0,i}(x)} \right| = \inf_{x \in I_n} \left| 1 + \left(\frac{\hat{g}_0(x) - \frac{1}{n} \sum_{i=1}^n g_{0,i}(x)}{\frac{1}{n} \sum_{i=1}^n g_{0,i}(x)} \right) \right|$$

$$\geq 1 - \frac{\sup_{x \in I_n} \left| \hat{g}_0(x) - \frac{1}{n} \sum_{i=1}^n g_{0,i}(x) \right|}{\inf_{x \in I_n} \left| \frac{1}{n} \sum_{i=1}^n f_{X_{i-1}}(x) \right|} = 1 + O_P \left(c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} q_n^f + c_n^2 q_n^f \right) \quad \text{(B.7)}$$

by (B.5), (B.6) and assumption (X) resp. (X').

Now for $\hat{m} = \hat{g}_1/\hat{g}_0$ from (2.3) we obtain

$$\sup_{x \in I_n} |\hat{m}(x) - m(x)|$$

$$= \sup_{x \in I_n} \left| \frac{\hat{g}_1(x) - \frac{1}{n} \sum_{i=1}^n g_{1,i}(x) + m(x) \left(\frac{1}{n} \sum_{i=1}^n g_{0,i}(x) - \hat{g}_0(x) \right)}{\hat{g}_0(x)} \right|$$

$$= \sup_{x \in I_n} \left| \frac{\frac{1}{\frac{1}{n} \sum_{i=1}^n f_{X_{i-1}}(x)} \left(\hat{g}_1(x) - \frac{1}{n} \sum_{i=1}^n g_{1,i}(x) + m(x) \left(\frac{1}{n} \sum_{i=1}^n g_{0,i}(x) - \hat{g}_0(x) \right) \right)}{\frac{\hat{g}_0(x)}{\frac{1}{n} \sum_{i=1}^n g_{0,i}(x)}} \right|$$

$$\leq \frac{\frac{1}{\inf_{x \in I_n} \frac{1}{n} \sum_{i=1}^n f_{X_{i-1}}(x)}}{\inf_{x \in I_n} \left| \frac{\hat{g}_0(x)}{\frac{1}{n} \sum_{i=1}^n g_{0,i}(x)} \right|} \left(\sup_{x \in I_n} \left| \hat{g}_1(x) - \frac{1}{n} \sum_{i=1}^n g_{1,i}(x) \right| + \sup_{x \in I_n} |m(x)| \sup_{x \in I_n} \left| \frac{1}{n} \sum_{i=1}^n g_{0,i}(x) - \hat{g}_0(x) \right| \right)$$

$$= O_P \left(c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} q_n q_n^f + c_n^2 q_n q_n^f \right)$$

28

by (B.5), (B.6), (B.7) and assumption (X) resp. (X'). The first assertion in (i) now directly follows from $(\inf_{x \in I_n} |\sigma(x)|)^{-1} = O(q_n^\sigma)$ (assumption (M)).

Differentiating it is easy to see that from (B.5) and (B.6) it follows that

$$
\sup_{x \in I_n} \left| \frac{\partial^\nu}{\partial x^\nu} \left(\frac{\hat{m}(x) - m(x)}{\sigma(x)} \right) \right|
$$

$$
= O\left((q_n^\sigma)^{\nu+1} \right) \sum_{j=0}^{\nu} O\left((q_n^f)^{j+1} \right) \sum_{l=0}^{1} O\left(q_n^{1-l} \right) \cdot O\left(\sup_{x \in I_n} \left| \frac{\partial^j}{\partial x^j} \left(\hat{g}_l(x) - \frac{1}{n} \sum_{i=1}^{n} g_{l,i}(x) \right) \right| \right)
$$

$$
= O_P\left(\left(c_n^{-\frac{1}{2}-\nu} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + c_n^2 \right) q_n (q_n^f q_n^\sigma)^{\nu+1} \right) = o_P(1),
$$

where the last equality only holds for $\nu = 0, 1$ by our bandwidth conditions. The first assertion of (ii) follows.

Finally, note that by considering the cases $|x - y| \leq c_n$ and $|x - y| > c_n$ we have

$$
\sup_{x,y \in I_n, x \neq y} \frac{\left| \frac{\partial}{\partial x} \left(\frac{\hat{m}(x) - m(x)}{\sigma(x)} \right) - \frac{\partial}{\partial y} \left(\frac{\hat{m}(y) - m(y)}{\sigma(y)} \right) \right|}{|y - x|^\delta}
$$

$$
\leq 2 \cdot \sup_{x \in I_n} \left| \frac{\partial}{\partial x} \left(\frac{\hat{m}(x) - m(x)}{\sigma(x)} \right) \right| c_n^{-\delta} + \sup_{x \in I_n} \left| \frac{\partial^2}{\partial x^2} \left(\frac{\hat{m}(x) - m(x)}{\sigma(x)} \right) \right| \sup_{x,y \in I_n, 0 < |x-y| \leq c_n} |x - y|^{1-\delta}
$$

$$
= O_P\left(\left(c_n^{-\frac{3}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + c_n^2 \right) q_n (q_n^f q_n^\sigma)^2 \right) c_n^{-\delta} + O_P\left(\left(c_n^{-\frac{5}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + c_n^2 \right) q_n (q_n^f q_n^\sigma)^3 \right) c_n^{(1-\delta)}
$$

$$
= o_P(1)
$$

and the first assertion of (iii) follows. □

Proof of Lemma B.3. (i). For

$$
\hat{d}_n(x) = w_n(x) \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \left(\frac{\frac{1}{c_n} K\left(\frac{x - X_{i-1}}{c_n} \right) - \frac{1}{nc_n} \sum_{k=1}^{n} K\left(\frac{x - X_{k-1}}{c_n} \right)}{\frac{1}{nc_n} \sum_{k=1}^{n} K\left(\frac{x - X_{k-1}}{c_n} \right)} \right)
$$

we have

$$
\int \hat{d}_n(x) f_{X_0}(x)\, dx = \frac{1}{n} \sum_{i=1}^{n} Z_i \int w_n(x) \left(\frac{1}{c_n} K\left(\frac{x - X_{i-1}}{c_n} \right) - f_{X_0}(x) \right) dx + o_P(\frac{1}{\sqrt{n}})
$$

$$
= o_P(\frac{1}{\sqrt{n}}),
$$

where last equality follows by a calculation of the variance and Chebyshev's inequality. The first equality can be derived by using

$$
\sup_{x \in I_n} \left| \frac{1}{nc_n} \sum_{k=1}^{n} K\left(\frac{x - X_{k-1}}{c_n} \right) - f_{X_0}(x) \right| = O_P\left(c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + c_n^2 \right), \tag{B.8}
$$

29

which follows from the proof of Lemma B.2 (note that with the notations used there, $(nc_n)^{-1}\sum_{k=1}^n K((x-X_{k-1})/c_n) = \hat{g}_0(x)$, $f_{X_0}(x) = n^{-1}\sum_{i=1}^n g_{0,i}(x))$ and results from Lemma B.2 (i) for the AR-model with $m \equiv 0$, $\sigma \equiv 1$.

Now assertion (i) is equivalent to

$$\sup_{s\in[0,1]} \left| \frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} \tilde{d}_n(X_{j-1}) \right| = o_P(\frac{1}{\sqrt{n}}).$$

with centered functions $\tilde{d}_n = \hat{d}_n - \int \hat{d}_n(x) f_{X_0}(x)\,dx$. By arguments similar to those in the proof of Lemma B.2 and by results from Lemma B.2 for $m \equiv 0$, $\sigma \equiv 1$ we have $P(\tilde{d}_n \in \mathcal{D}_n) \to 1$ for $n \to \infty$ for the function classes

$$\mathcal{D}_n = \left\{ d : I_n \to \mathbb{R} \,\middle|\, \max\left(\sup_{x\in I_n} |d(x)|, \sup_{x\in I_n} |d'(x)|\right) + \sup_{x,y\in I_n, x\neq y} \frac{|d'(x) - d'(y)|}{|x - y|^\delta} \leq 1, \right.$$
$$\left. \sup_{x\in I_n} |d(x)| \leq z_n, \int d(y) f_{X_0}(y) dy = 0 \right\}$$

with $z_n = c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} q_n^f (\log n)^{\frac{1}{2}}$. Thus it remains to show that

$$\sup_{s\in[0,1]} \sup_{d\in\mathcal{D}_n} \left| \frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} d(X_{j-1}) \right| = o_P(\frac{1}{\sqrt{n}}). \tag{B.9}$$

To this end let $\bar{\epsilon}_n = n^{-\frac{1}{2}}(\log n)^{-1}$. It follows from Theorem 2.7.1 in van der Vaart & Wellner (1996) that $M_{n2} \leq \exp(2^{\frac{1}{1+\delta}} \bar{K}(2 + b_n - a_n)\bar{\varepsilon}_n^{-\frac{1}{1+\delta}})$ balls of radius $\bar{\epsilon}_n$ with respect to the supremum norm $||\cdot||_{I_n}$ on the interval I_n are needed to cover $\mathcal{D}_n$. Here the constant $\bar{K}$ only depends on δ. Let $d_1, \ldots, d_{M_{n2}}$ denote centers of those balls. We may assume that those functions are elements of $\mathcal{D}_n$, too (see Pollard (1990), p. 10). Further let $0 = s_1 < \ldots < s_{M_{n1}} = 1$ segment $[0,1]$ in intervals of length $\leq \bar{\epsilon}_n/z_n$ such that $M_{n1} \leq \frac{z_n}{\bar{\epsilon}_n}$. Then it can be

30

shown that

$$
\sup_{s\in[0,1]} \sup_{d\in\mathcal{D}_n} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns\rfloor} d(X_{j-1}) \right|
$$

$$
\leq \max_{1\leq h\leq M_{n1}} \max_{1\leq k\leq M_{n2}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h\rfloor} d_k(X_{j-1}) \right|
$$

$$
+ \max_{1\leq h\leq M_{n1}} \sup_{s\in[0,1] \text{ mit } |s-s_h|\leq \bar{\epsilon}_n/z_n} \sup_{d\in\mathcal{D}_n} \left| \frac{1}{n} \sum_{j=1}^{n} d(X_{j-1}) \left(I\left\{ \frac{j}{n} \leq s \right\} - I\left\{ \frac{j}{n} \leq s_h \right\} \right) \right|
$$

$$
+ \max_{h} \max_{1\leq k\leq M_{n2}} \sup_{d\in\mathcal{D}_n \text{ mit } \|d-d_k\|_{I_n}\leq \bar{\epsilon}_n} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h\rfloor} (d(X_{j-1}) - d_k(X_{j-1})) \right|
$$

$$
\leq \max_{1\leq h\leq M_{n1}} \max_{1\leq k\leq M_{n2}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h\rfloor} d_k(X_{j-1}) \right| + o\left(\frac{1}{\sqrt{n}} \right).
$$

By an application of Liebscher's (1996) Theorem 2.1 to random variables $Y_i = d_k(X_{i-1})I\left\{ \frac{i}{n} \leq s_h \right\}$ (for k, h fixed) one can show the existence of some constant $\bar{M}$ such that

$$
P\left(\max_{1\leq h\leq M_{n1}} \max_{1\leq k\leq M_{n2}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h\rfloor} d_k(X_{j-1}) \right| > \bar{M}\bar{\epsilon}_n \right)
$$

$$
\leq \sum_{h=1}^{M_{n1}} \sum_{k=1}^{M_{n2}} P\left(\left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h\rfloor} d_k(X_{j-1}) \right| > \bar{M}\bar{\epsilon}_n \right)
$$

$$
\leq M_{n1}M_{n2} \left(4\exp\left(-\frac{n^2 \bar{M}^2 \bar{\epsilon}_n^2}{64n\lfloor n\bar{\epsilon}_n c_n^{\frac{1}{2}}\rfloor z_n^2 + \frac{8}{3}n\bar{M}\bar{\epsilon}_n\lfloor n\bar{\epsilon}_n c_n^{\frac{1}{2}}\rfloor z_n} \right) + 4\frac{n}{\lfloor n\bar{\epsilon}_n c_n^{\frac{1}{2}}\rfloor}\alpha(\lfloor n\bar{\epsilon}_n c_n^{\frac{1}{2}}\rfloor) \right)
$$

$$
= o(1).
$$

Details are omitted for the sake of brevity. From this the rate $O_P(\bar{\epsilon}_n) = o_P(n^{-1/2})$ follows for (B.9).

(ii). We only describe the main steps of this proof. The random denominator can be replaced by the true density f_{X_0} due to (B.8). Now define

$$
\hat{d}_n(x) = w_n(x)\frac{1}{nc_n} \sum_{i=1}^{n} K\left(\frac{x - X_{i-1}}{c_n} \right) \varepsilon_i \frac{\sigma(X_{i-1}) - \sigma(x)}{\sigma(x)f_{X_0}(x)}.
$$

Then, $\int \hat{d}_n(x)f_{X_0}(x)\,dx = o_P(n^{-1/2})$ can be shown by Chebyshev's inequality and for $\tilde{d}_n = \hat{d}_n - \int \hat{d}_n(x)f_{X_0}(x)\,dx$ the assertion

$$
\sup_{s\in[0,1]} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns\rfloor} \tilde{d}_n(X_{j-1}) \right| = o_P(\frac{1}{\sqrt{n}}).
$$

31

is shown analogously to the proof of (i).

(iii). The assertion can be proved by the same methods. $\qquad\square$

Proof of Lemma B.4. Let $\hat{d}_{n1} = (\hat{m} - m)/\sigma$ and $\hat{d}_{n2} = \hat{\sigma}/\sigma$. Now the assumption of the lemma is equivalent to

$$\sup_{s\in[0,1]} \sup_{t\in\mathbb{R}} |H_n(s,t,\hat{d}_{n1},\hat{d}_{n2}) - H_n(s,t,0,1)| = o_P(\frac{1}{\sqrt{n}})$$

with

$$H_n(s,t,d_1,d_2) = \frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} w_{nj}\left(I\{\epsilon_j \le t\cdot d_2(X_{j-1}) + d_1(X_{j-1})\} - F_{\varepsilon_j}(td_2(X_{j-1}) + d_1(X_{j-1}))\right).$$

For the proof we may assume that $\sup_{x\in I_n}|\hat{d}_{n1}(x)| < 1$, $\inf_{x\in I_n}\hat{d}_{n2}(x) > \frac{1}{2}$ and $|\varepsilon_j| \le \sqrt{n}\log n$ for all $j = 1,\ldots,n$ because $\sup_{x\in I_n}|\hat{d}_{n1}(x)| = o_P(1)$, $\sup_{x\in I_n}|\hat{d}_{n2}(x) - 1| = o_P(1)$ by Lemma B.2, and further

$$P\left(\max_{1\le j\le n}|\varepsilon_j| > \sqrt{n}\log n\right) \le \frac{1}{n\log n}\sum_{j=1}^{n}E\left[\varepsilon_j^2 I\{\varepsilon_j^2 > n(\log n)^2\}\right] \le \frac{1}{\log n} = o(1)$$

by the model assumption $E[\varepsilon_j^2] = 1$ for all $j \in \mathbb{Z}$.

Note that $I\{\varepsilon_j \le t\cdot\hat{d}_{n2}(X_{j-1}) + \hat{d}_{n1}(X_{j-1})\} - I\{\varepsilon_j \le t\} = 0$ for all t such that $|t| > \sqrt{n}\log n$, for all $j = 1,\ldots,n$. Thus, by some simple estimations,

$$\sup_{s\in[0,1]} \sup_{|t|>\sqrt{n}\log n} |H_n(s,t,\hat{d}_{n1},\hat{d}_{n2}) - H_n(s,t,0,1)|$$

$$\le \sup_{s\in[0,1]} \sup_{|t|>\sqrt{n}\log n} \frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} w_{nj}\left|-F_{\varepsilon_j}(t\cdot\hat{d}_{n2}(X_{j-1}) + \hat{d}_{n1}(X_{j-1})) + F_{\varepsilon_j}(t)\right|$$

$$\le \frac{1}{n}\sum_{j=1}^{n}\left(1 - F_{\varepsilon_j}\left(\frac{\sqrt{n}\log n}{2} - 1\right)\right) + \frac{1}{n}\sum_{j=1}^{n}F_{\varepsilon_j}\left(-\frac{\sqrt{n}\log n}{2} + 1\right)$$

$$= O\left(\frac{1}{n(\log n)^2}\right) = o\left(\frac{1}{\sqrt{n}}\right),$$

where in the last line we have applied that for $t > 1$

$$\frac{1}{n}\sum_{j=1}^{n}(1 - F_{\varepsilon_j}(t)) = \frac{1}{n}\sum_{j=1}^{n}P(\varepsilon_j > t) \le \frac{1}{t^2}\frac{1}{n}\sum_{j=1}^{n}E[\varepsilon_j^2 I\{\varepsilon_j^2 > t^2\}] = O\left(\frac{1}{t^2}\right) \quad \text{(B.10)}$$

by the model assumption $E[\varepsilon_j^2] = 1$ and analogously for $t < -1$, $\frac{1}{n}\sum_{j=1}^{n}F_{\varepsilon_j}(t) = O(1/t^2)$.

For the remainder of the proof we therefore only need to consider $|t| \leq \sqrt{n} \log n$. Define sequences of function classes by

$$\mathcal{D}_{1,n} = \left\{ d : I_n \to \mathbb{R} \,\Big|\, \max\{\sup_{x \in I_n} |d(x)|, \sup_{x \in I_n} |d'(x)|\} + \sup_{x,y \in I_n, x \neq y} \frac{|d'(x) - d'(y)|}{|y - x|^\delta} \leq 1, \right.$$

$$\left. \sup_{x \in I_n} |d(x)| \leq z_{n1} \log n \right\} \text{ for } z_{n1} = \left(c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + c_n^2 \right) q_n q_n^f q_n^\sigma$$

$$\mathcal{D}_{2,n} = \left\{ d : I_n \to \mathbb{R} \,\Big|\, \max\{\sup_{x \in I_n} |d(x)|, \sup_{x \in I_n} |d'(x)|\} + \sup_{x,y \in I_n, x \neq y} \frac{|d'(x) - d'(y)|}{|y - x|^\delta} \leq 2, \right.$$

$$\left. \inf_{x \in I_n} d(x) \geq \frac{1}{2}, \sup_{x \in I_n} |d(x) - 1| \leq z_{n2} \log n \right\} \text{ for } z_{n2} = \left(c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + c_n^2 \right) (q_n q_n^f q_n^\sigma)^2.$$

Then by Lemma B.2, $P(\hat{d}_{n1} \in \mathcal{D}_{1,n}) \to 1$, $P(\hat{d}_{n2} \in \mathcal{D}_{2,n}) \to 1$ and it remains to show that

$$\sup_{\substack{s \in [0,1], |t| \leq \sqrt{n} \log n, \\ d_1 \in \mathcal{D}_{1,n}, d_2 \in \mathcal{D}_{2,n}}} |H_n(s,t,d_1,d_2) - H_n(s,t,0,1)| = o_p\left(\frac{1}{\sqrt{n}}\right).$$

To this end we apply covering arguments. Let $\bar{\epsilon}_n = \min\{\frac{1}{8}, n^{-\frac{1}{2}}(\log n)^{-1}\}$. The $\bar{\epsilon}_n$-covering numbers of both function classes with respect to the supremum norm on I_n can be bounded by $M_n \leq \exp(c(2 + b_n - a_n)\bar{\epsilon}_n^{-1/(1+\delta)})$, see Theorem 2.7.1 by van der Vaart & Wellner (1996). Let $d_{11}, \ldots, d_{1M_n}$ and $d_{21}, \ldots, d_{2M_n}$, respectively, denote the corresponding centers of covering balls. Note that then $\sup_{x \in I_n} |d_{1k}(x)| \leq 1 + \bar{\epsilon}_n$ and $d_{2l}(x) \in [\frac{1}{2} - \bar{\epsilon}_n, 2 + \bar{\epsilon}_n]$ for all $x \in I_n$. Let further the intervals $[0,1]$ and $[-\sqrt{n} \log n, \sqrt{n} \log n]$ be segmented by points $0 = s_1 < \ldots < s_{M_{ns}} = 1$ and $-\sqrt{n} \log n = t_1 < \ldots < t_{M_{nt}} = \sqrt{n} \log n$, respectively, in segments of length $\leq \bar{\epsilon}_n$ such that the number of points are bounded by $M_{ns} \leq 1/\bar{\epsilon}_n$ and $M_{nt} \leq 2\sqrt{n} \log n/\bar{\epsilon}_n$. Let $|| \cdot ||_{I_n}$ denote the supremum norm on I_n. Then

$$\sup_{\substack{s \in [0,1], |t| \leq \sqrt{n} \log n, \\ d_1 \in \mathcal{D}_{1,n}, d_2 \in \mathcal{D}_{2,n}}} |H_n(s,t,d_1,d_2) - H_n(s,t,0,1)|$$

$$\leq \max_{h,i,k,l} |H_n(s_h, t_i, d_{1k}, d_{2l}) - H_n(s_h, t_i, 0, 1)|$$

$$+ \max_h \sup_{\substack{|s - s_h| \leq \bar{\epsilon}_n, |t| \leq \sqrt{n} \log n, \\ d_1 \in \mathcal{D}_{1,n}, d_2 \in \mathcal{D}_{2,n}}} |H_n(s, t, d_1, d_2) - H_n(s_h, t, d_1, d_2)| \tag{B.11}$$

$$+ \max_{h,i,k,l} \sup_{\substack{|t - t_i| \leq \bar{\epsilon}_n, \\ \|d_1 - d_{1k}\|_{I_n} \leq \bar{\epsilon}_n, \|d_2 - d_{2l}\|_{I_n} \leq \bar{\epsilon}_n}} |H_n(s_h, t, d_1, d_2) - H_n(s_h, t_i, d_{1k}, d_{2l})| \tag{B.12}$$

$$+ \max_h \sup_{|s - s_h| \leq \bar{\epsilon}_n, |t| \leq \sqrt{n} \log n} |H_n(s_h, t, 0, 1) - H_n(s, t, 0, 1)| \tag{B.13}$$

$$+ \max_{h,i} \sup_{|t - t_i| \leq \bar{\epsilon}_n} |H_n(s_h, t_i, 0, 1) - H_n(s_h, t, 0, 1)|, \tag{B.14}$$

33

where the maximum is always with respect to $h \in \{1, \ldots, M_{ns}\}$, $i \in \{1, \ldots, M_{nt}\}$, $k, l \in \{1, \ldots, M_n\}$.

To further bound the term (B.12) first consider fixed $h \in \{1, \ldots, M_{ns}\}$, $i \in \{1, \ldots, M_{nt}\}$, $k, l \in \{1, \ldots, M_n\}$ such that $t_i \geq \bar{\epsilon}_n$ (the other case is treated analogously). Then

$$
\sup_{\substack{|t - t_i| \leq \bar{\epsilon}_n, \\ \|d_1 - d_{1k}\|_{I_n} \leq \bar{\epsilon}_n, \|d_2 - d_{2l}\|_{I_n} \leq \bar{\epsilon}_n}} |H_n(s_h, t, d_1, d_2) - H_n(s_h, t_i, d_{1k}, d_{2l})|
$$

$$
\leq \sup_{\substack{|t - t_i| \leq \bar{\epsilon}_n, \\ \|d_1 - d_{1k}\|_{I_n} \leq \bar{\epsilon}_n, \|d_2 - d_{2l}\|_{I_n} \leq \bar{\epsilon}_n}} \frac{1}{n} \sum_{j=1}^{n} w_{nj} |I\{\varepsilon_j \leq td_2(X_{j-1}) + d_1(X_{j-1})\} - I\{\varepsilon_j \leq td_{2l}(X_{j-1}) + d_{1k}(X_{j-1})\}|
$$

$$
+ \sup_{\substack{|t - t_i| \leq \bar{\epsilon}_n, \\ \|d_1 - d_{1k}\|_{I_n} \leq \bar{\epsilon}_n, \|d_2 - d_{2l}\|_{I_n} \leq \bar{\epsilon}_n}} \frac{1}{n} \sum_{j=1}^{n} w_{nj} |F_{\varepsilon_j}(td_{2l}(X_{j-1}) + d_{1k}(X_{j-1})) - F_{\varepsilon_j}(td_2(X_{j-1}) + d_1(X_{j-1}))|
$$

$$
\leq H_n(1, t_i + \bar{\epsilon}_n, d_{1k} + \bar{\epsilon}_n, d_{2l} + \bar{\epsilon}_n) - H_n(1, t_i - \bar{\epsilon}_n, d_{1k} - \bar{\epsilon}_n, d_{2l} - \bar{\epsilon}_n)
$$

$$
+ \frac{2}{n} \sum_{j=1}^{n} w_{nj} \Bigg(F_{\varepsilon_j}((t_i + \bar{\epsilon}_n)(d_{2l}(X_{j-1}) + \bar{\epsilon}_n) + d_{1k}(X_{j-1}) + \bar{\epsilon}_n)
$$

$$
- F_{\varepsilon_j}((t_i - \bar{\epsilon}_n)(d_{2l}(X_{j-1}) - \bar{\epsilon}_n) + d_{1k}(X_{j-1}) - \bar{\epsilon}_n) \Bigg)
$$

$$
= H_n(1, t_i + \bar{\epsilon}_n, d_{1k} + \bar{\epsilon}_n, d_{2l} + \bar{\epsilon}_n) - H_n(1, t_i - \bar{\epsilon}_n, d_{1k} - \bar{\epsilon}_n, d_{2l} - \bar{\epsilon}_n) + o\left(\frac{1}{\sqrt{n}}\right),
$$

where the last step follows from the mean value theorem, assumption (F) resp. (F') and $\bar{\epsilon}_n = o(1/\sqrt{n})$. Similarly for (B.14) we obtain

$$
\max_{h,i} \sup_{|t - t_i| \leq \bar{\epsilon}_n} |H_n(s_h, t_i, 0, 1) - H_n(s_h, t, 0, 1)|
$$

$$
\leq \max_i |H_n(1, t_i + \bar{\epsilon}_n, 0, 1) - H_n(1, t_i - \bar{\epsilon}_n, 0, 1)| + o\left(\frac{1}{\sqrt{n}}\right). \tag{B.15}
$$

For (B.11) we have

$$
\max_h \sup_{\substack{|s - s_h| \leq \bar{\epsilon}_n, |t| \leq \sqrt{n} \log n, \\ d_1 \in \mathcal{D}_{1,n}, d_2 \in \mathcal{D}_{2,n}}} |H_n(s, t, d_1, d_2) - H_n(s_h, t, d_1, d_2)|
$$

$$
\leq \max_h \sup_{|s - s_h| \leq \bar{\epsilon}_n} \frac{1}{n} \sum_{j=1}^{n} \left| I\left\{\frac{j}{n} \leq s\right\} - I\left\{\frac{j}{n} \leq s_h\right\} \right|
$$

$$
\leq \max_h \sup_{|s - s_h| \leq \bar{\epsilon}_n} \left(|s - s_h| + \frac{1}{n}\right) \leq \bar{\epsilon}_n + \frac{1}{n} = o\left(\frac{1}{\sqrt{n}}\right)
$$

34

and analogously for (B.13) the same rate $o(1/\sqrt{n})$. Altogether we have shown that

$$\sup_{\substack{s\in[0,1],|t|\leq\sqrt{n}\log n,\\ d_1\in\mathcal{D}_{1,n},d_2\in\mathcal{D}_{2,n}}}|H_n(s,t,d_1,d_2)-H_n(s,t,0,1)|$$

$$\leq \max_{h,i,k,l}|H_n(s_h,t_i,d_{1k},d_{2l})-H_n(s_h,t_i,0,1)| \tag{B.16}$$

$$+\max_{i,k,l}|H_n(1,t_i+\bar{\epsilon}_n,d_{1k}+\bar{\epsilon}_n,d_{2l}+\bar{\epsilon}_n)-H_n(1,t_i-\bar{\epsilon}_n,d_{1k}-\bar{\epsilon}_n,d_{2l}-\bar{\epsilon}_n)|$$

$$+\max_{i}|H_n(1,t_i+\bar{\epsilon}_n,0,1)-H_n(1,t_i-\bar{\epsilon}_n,0,1)|+o(\frac{1}{\sqrt{n}}).$$

To conclude the proof we exemplarily consider term (B.16); the other terms are treated analogously. For all $\eta>0$ we have

$$P\left(\sqrt{n}\max_{h,i,k,l}|H_n(s_h,t_i,d_{1k},d_{2l})-H_n(s_h,t_i,0,1)|>\eta\right)$$

$$\leq \sum_{h=1}^{M_{ns}}\sum_{i=1}^{M_{nt}}\sum_{k=1}^{M_n}\sum_{l=1}^{M_n}P\left(\sqrt{n}\,|H_n(s_h,t_i,d_{1k},d_{2l})-H_n(s_h,t_i,0,1)|>\eta\right)$$

$$\leq M_{ns}\sum_{i=1}^{M_{nt}}\sum_{k=1}^{M_n}\sum_{l=1}^{M_n}\left(2\exp\left(-\frac{\eta^2 n}{4\eta\sqrt{n}+2nA_{i,k,l}(n)}\right)\right) \tag{B.17}$$

by an application of Theorem 1.6 by Freedman (1975). To see this define (for h,i,k,l fixed)

$$Y_j \;=\; w_{nj}I\left\{\frac{j}{n}\leq s_h\right\}\big(I\{\varepsilon_j\leq t_id_{2l}(X_{j-1})+d_{1k}(X_{j-1})\}-I\{\varepsilon_j\leq t_i\}$$

$$-F_{\varepsilon_j}(t_id_{2l}(X_{j-1})+d_{1k}(X_{j-1}))+F_{\varepsilon_j}(t_i)\big)\,,$$

and note that $|Y_j|\leq 2$, $E[Y_j|X_0,\ldots,X_{j-1}]=0=E[-Y_j|X_0,\ldots,X_{j-1}]$ as well as

$$\sum_{j=1}^{n}E\left[Y_j^2|X_0,\ldots,X_{j-1}\right] \;\leq\; nA_{i,k,l}(n) \;=\; \sum_{j=1}^{n}\sup_{x\in I_n}|F_{\varepsilon_j}(t_id_{2l}(x)+d_{1k}(x))-F_{\varepsilon_j}(t_i)|$$

$$\leq\; Cn(z_{n1}\log n+z_{n2}\log n+\bar{\epsilon}_n)$$

by an application of the mean value theorem, the definition of $\mathcal{D}_{1,n}$, $\mathcal{D}_{2,n}$ and assumption (F) resp. (F'), for some constant C independent of i,k,l. Hence, inserting the bounds on M_n, M_{ns} and M_{nt}, (B.16) can be bounded by

$$2\exp\left(\log\left(\frac{\sqrt{n}\log n}{\bar{\epsilon}_n^2}\right)+2c(2+b_n-a_n)\bar{\varepsilon}_n^{-\frac{1}{1+\delta}}\right.$$

$$\left.-\frac{n\eta^2}{4\eta\sqrt{n}+2Cn(z_{n1}\log n+z_{n2}\log n+\bar{\epsilon}_n)}\right)=o(1),$$

which follows from bandwidth condition (3.1). □

Proof of Lemma B.5. Analogous to the proof of Lemma B.4 we may assume that $|\varepsilon_j| \leq \sqrt{n}\log n$ for all $j = 1, \ldots, n$. Then

$$\sup_{\substack{s \in [0,1], \\ |t| > \sqrt{n}\log n}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} (w_n(X_{j-1}) - 1)\left(I\{\varepsilon_j \leq t\} - F(t)\right) \right|$$

$$\leq \sup_{\substack{s \in [0,1], \\ t > \sqrt{n}\log n}} \left| \frac{1}{n} \sum_{j=1}^{n} I\left\{ \frac{j}{n} \leq s \right\} (w_n(X_{j-1}) - 1)(1 - F(t)) \right|$$

$$+ \sup_{\substack{s \in [0,1], \\ t < -\sqrt{n}\log n}} \left| \frac{1}{n} \sum_{j=1}^{n} I\left\{ \frac{j}{n} \leq s \right\} (w_n(X_{j-1}) - 1)(0 - F(t)) \right|$$

$$\leq \frac{1}{n} \sum_{j=1}^{n} \left(1 - F(\sqrt{n}\log n)\right) + \frac{1}{n} \sum_{j=1}^{n} F(-\sqrt{n}\log n)$$

$$= O\left(\frac{1}{n(\log n)^2}\right) = o\left(\frac{1}{\sqrt{n}}\right),$$

where the second last equality follows from (B.10).

Now let $\bar{\epsilon}_n = n^{-\frac{1}{2}}(\log n)^{-1}$ and let $0 = s_1 < \ldots < s_{M_{n,1}} = 1$ be such that $s_h - s_{h-1} \leq \bar{\epsilon}_n$ for all $i = 2, \ldots, M_{n,1}$ and $M_{n,1} \leq 1/\bar{\epsilon}_n$. Further let $-\sqrt{n}\log n = t_1 < \ldots < t_{M_{n,2}} = \sqrt{n}\log n$

36

be such that $t_i - t_{i-1} \leq \bar{\epsilon}_n$ for all $i = 2, \ldots, M_{n,2}$ and $M_{n,2} \leq 2\sqrt{n} \log n / \bar{\epsilon}_n$. Then we have

$$\sup_{\substack{s \in [0,1], \\ t \in [-\sqrt{n} \log n, \sqrt{n} \log n]}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t\} - F(t)) \right|$$

$$\leq \max_{\substack{1 \leq h \leq M_{n,1}, \\ 1 \leq i \leq M_{n,2}}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t_i\} - F(t_i)) \right|$$

$$+ \max_{1 \leq h \leq M_{n,1}} \sup_{\substack{|s-s_h| \leq \bar{\epsilon}_n, \\ t \in [-\sqrt{n} \log n, \sqrt{n} \log n]}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t\} - F(t)) \right.$$

$$\left. - \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t\} - F(t)) \right| \quad \text{(B.18)}$$

$$+ \max_{\substack{1 \leq h \leq M_{n,1}, \\ 1 \leq i \leq M_{n,2}}} \sup_{|t-t_i| \leq \bar{\epsilon}_n} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t\} - F(t)) \right.$$

$$\left. - \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t_i\} - F(t_i)) \right| \quad \text{(B.19)}$$

$$\leq \max_{\substack{1 \leq h \leq M_{n,1}, \\ 1 \leq i \leq M_{n,2}}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) (I\{\varepsilon_j \leq t_i\} - F(t_i)) \right| \quad \text{(B.20)}$$

$$+ o\left(\frac{1}{\sqrt{n}}\right)$$

$$+ \max_{1 \leq i \leq M_{n,2}} \left| \frac{1}{n} \sum_{j=1}^{n} (w_n(X_{j-1} - 1))(I\{\varepsilon_j \leq t_i + \bar{\epsilon}_n\} - F(t_i + \bar{\epsilon}_n)) \right| \quad \text{(B.21)}$$

$$+ \max_{1 \leq i \leq M_{n,2}} \left| \frac{1}{n} \sum_{j=1}^{n} (w_n(X_{j-1} - 1))(I\{\varepsilon_j \leq t_i - \bar{\epsilon}_n\} - F(t_i - \bar{\epsilon}_n)) \right| \quad \text{(B.22)}$$

$$+ 2 \max_{1 \leq i \leq M_{n,2}} \frac{1}{n} \sum_{j=1}^{n} (F(t_i + \bar{\epsilon}_n) - F(t_i - \bar{\epsilon}_n)) . \quad \text{(B.23)}$$

To obtain the last inequality it can be shown analogously to the treatment of (B.11) in the proof of Lemma B.4 that (B.18) is of order $O(\bar{\epsilon}_n) = o(1/\sqrt{n})$. Further the bounding of (B.19) by the sum of (B.21), (B.22) and (B.23) is straightforward by using monotonicity of indicator and distribution functions.

Now by the mean value theorem and assumption (F) it follows that (B.23) is of order $O(\bar{\epsilon}_n) = o(1/\sqrt{n})$. The remaining terms (B.20), (B.21), (B.22) are treated in the same way

and we will only consider (B.20) in what follows. For this term we have for each $\eta > 0$ that

$$
P\left(\sqrt{n} \max_{\substack{1 \leq h \leq M_{n,1}, \\ 1 \leq i \leq M_{n,2}}} \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) \left(I\{\varepsilon_j \leq t_i\} - F(t_i)\right) \right| > \eta\right)
$$

$$
\leq \sum_{h=1}^{M_{n,1}} \sum_{i=1}^{M_{n,2}} P\left(\sqrt{n} \cdot \left| \frac{1}{n} \sum_{j=1}^{\lfloor ns_h \rfloor} (w_n(X_{j-1}) - 1) \left(I\{\varepsilon_j \leq t_i\} - F(t_i)\right) \right| > \eta\right)
$$

$$
\leq M_{n,1} M_{n,2} \left(4 \exp\left(-\frac{n\eta^2}{64 n \omega_n + \frac{8}{3}\sqrt{n}\eta \lfloor n^{\frac{1}{2}}(\log n)^{-2} \rfloor} \right) + 4 \frac{n}{\lfloor n^{\frac{1}{2}}(\log n)^{-2} \rfloor} \alpha(\lfloor n^{\frac{1}{2}}(\log n)^{-2} \rfloor) \right)
$$

$$
= o(1)
$$

(where $\alpha(\cdot)$ denotes the α-mixing coefficient) by an application of Theorem 2.1 by Liebscher (1996) and the bandwidth conditions. Details are omitted for the sake of brevity, but note that

$$
\omega_n = \frac{1}{\lfloor n^{\frac{1}{2}}(\log n)^{-2} \rfloor} \max_{0 \leq S \leq n - \lfloor n^{\frac{1}{2}}(\log n)^{-2} \rfloor} \sum_{j=S+1}^{S + \lfloor n^{\frac{1}{2}}(\log n)^{-2} \rfloor} E\left[(w_n(X_{j-1}) - 1)^2\right]
$$

$$
\leq \int_{-\infty}^{a_n + \kappa} f_{X_0}(x)dx + \int_{b_n - \kappa}^{\infty} f_{X_0}(x)dx = o\left(\frac{1}{\log n}\right)
$$

by assumption (I). □

Proof of Lemma B.6. We only give arguments for the first statement. Similarly to the proof of Lemma B.5 one can show that

$$
\frac{1}{n} \sum_{j=1}^{\lfloor n\theta_0 \rfloor} w_{nj} \left(I\{\varepsilon_j \leq t\} - F(t)\right) = \frac{1}{n} \sum_{j=1}^{\lfloor n\theta_0 \rfloor} \left(I\{\varepsilon_j \leq t\} - F(t)\right) + o_P(1)
$$

(applying assumptions (I') and (X')). Then the assertion follows by standard arguments for the empirical distribution function of iid-data. □

References

Andreou, E. & Ghysels, E. (2009) *Structural Breaks in Financial Time Series.* T. G. Andersen (ed) et al, Handbook of Financial Time Series. Springer, Berlin, 839-870.

Bai, J. (1994). *Weak convergence of the sequential empirical processes of residuals in ARMA models.* Ann. Statist. 22, 2051-2061.

Boldin, M. V. (2002). *On sequential residual empirical processes in heteroscedastic time series.* Math. Methods Statist. 11, 453-464.

Csörgö, M., Horváth, L. & Szyszkowicz (1997). *Integral tests for suprema of Kiefer processes with application.* Statist. Decisions 15, 365-377.

Dette, H., Pardo-Fernández, J. C. & Van Keilegom, I. (2009). *Goodness-of-Fit Tests for Multiplicative Models with Dependent Data.* Scand. J. Statist. 36, 782-799.

Doukhan, P. (1994). *Mixing, Properties and Examples.* Springer, New York.

Fan, J. & Yao, Q. (2005). *Nonlinear Time Series.* Springer, New York.

Freedman, D. A. (1975). *On tail probabilities for martingals.* Ann. Probab. 3, 100-118.

Giraitis, L., Leipus, R. & Surgailis, D. (1996) *The change-point problem for dependent observations.* J. Statist. Plann. Inf. 53, 297-310.

Hansen, B. E. (2008). *Uniform Convergence Rates for Kernel Estimation with Dependent Data.* Econom. Theory 24, 726-748.

Hlávka, Z., Hušková, M., Kirch, C. & Meintanis, S. (2012). *Monitoring changes in the error distribution of autoregressive models based on Fourier methods.* to appear in Test.

Horváth, L., Kokoszka, P. & Teyssière, G. (2001). *Empirical process of the squared residuals of an ARCH sequence.* Ann. Statist. 29, 445-469.

Hušková, M. & Antoch, J. (2003). *Detection of structural changes in regression.* Tatra Mt. Math. Publ. 26, 201-215.

Hušková, M., Prášková, Z. & Steinebach, J. (2007). *On the detection of changes in autoregressive time series. I. Asymptotics.* J. Statist. Plann. Inference 137, 1243-1259.

Hušková, M., Kirch, C., Prášková, Z. & Steinebach, J. (2008). *On the detection of changes in autoregressive time series. II. Resampling.* J. Statist. Plann. Inference 138, 1697-1721.

Inoue, A. (2001). *Testing for distributional change in time series.* Economet. Theory 17, 156-187.

Kirch, C. & Tadjuidje Kamgaing, J. (2012). *Testing for parameter stability in nonlinear autoregressive models.* J. Time Ser. Anal., to appear.

Koul, H. L. (1996). *Asymptotics of some estimators and sequential residual empiricals in nonlinear time series.* Ann. Statist. 24, 380-404.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models (Second Edition).* Springer, New York.

Kreiß, J.-P. (1991). *Estimation of the distribution function of noise in stationary processes.* Metrika 38, 285-297.

Lee, S. & Na, S. (2004). *A nonparametric test for the change in the density function in strong mixing processes.* Statist. Prob. Letters 66, 1-25.

Liebscher, E. (1996). *Strong convergence of sums of α-mixing random variables with applications to density estimation.* Stochastic Processes and their Applications 65, 69-80.

Müller, U. U., Schick, A. & Wefelmeyer, W. (2009). *Estimating the innovation distribution in nonparametric autoregression.* Probab. Theory Relat. Fields 144, 53-77.

Neumeyer, N. & Van Keilegom, I. (2009). *Change-Point Tests for the Error Distribution in Nonparametric Regression.* Scand. J. Statist. 36, 518-541.

Picard, D. (1985). *Testing and estimating change-points in time series.* Adv. Appl. Probab. 17, 841-867.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics 2, Institute of Mathematical Statistics.

Selk, L. (2011). *Change-Point-Tests für die Innovationenverteilung in nichtparametrischen Autoregressionsmodellen auf Basis sequentieller empirischer Prozesse.* PhD thesis (in German), Universität Hamburg. http://ediss.sub.uni-hamburg.de/volltexte/2011/5338/

Shao, X. & Zhang, X. (2010). *Testing for Change Points in Time Series.* J. Amer. Statist. Assoc. 105, 1228-1240.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical Processes with Applications ot Statistics.* Wiley, New York.

Shumway, R. H. & Stoffer, D.S. (2006). *Time Series Analysis and Its Applications: With R Examples.* Springer, New York.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.